UCRL-JRNL-206139

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Insights into the evolution of Yersinia pestis through whole genome comparison with Yersinia pseudotuberculosis

P. Chain, E. Carniel, F. Larimer, J. Lamerdin, P. Stoutland, W. Regala, A. Georgescu, L. Vergez, M. Land, V. Motin, R. Brubaker, J. Fowler, J. Hinnebusch, M. Marceau, C. Medigue, M. Simonet, V. Chenal-Francisque, B. Souza, D. Dacheux, J. Elliott, A. Derbise, L. Hauser, E. Garcia

August 19, 2004

Proceedings of the National Academy of Sciences

UCRL-JRNL-206139

Insights into the evolution of *Yersinia pestis* through whole genome comparison with *Yersinia pseudotuberculosis*

P. S. G. Chain[1], E. Carniel[2], F. W. Larimer[3], J. Lamerdin[1], P. O. Stoutland[1], W. M Regala[1], A. M. Georgescu[1], L. M. Vergez[1], M. L. Land[3], V. L. Motin[1], R. R. Brubaker[4], J. Fowler[4], J. Hinnebusch[5], M. Marceau[6], C. Medigue[7], M. Simonet[6], V. Chenal-Francisque[2], B. Souza[1], D. Dacheux[2], J. M. Elliott[1], A. Derbise[2], L. J. Hauser[3], and E. Garcia[1*]

[1]*Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA*
[2]*Yersinia Research Unit, Institut Pasteur, 75724 Paris Cedex 15, France*
[3]*Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[4]*Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824, USA*
[5]*Rocky Mountain Laboratories, Hamilton, MT 59840, USA*
[6]*Inserm E0364 -Université de Lille 2-Institut Pasteur de Lille, F-59021 Lille, France*
[7]*Genoscope/CNRS-UMR 8030, 91006 Evry Cedex, France*

* Corresponding author
Emilio Garcia
LLNL
7000 East Ave. L-452
Livermore, CA
94550
garcia12@llnl.gov
Ph. 925-422-8002
Fax 925-422-2099

23 text pages (including 2 Fig. legends), 1 Table, 2 Figures, 8 Supplemental Tables
Words in abstract: 195; Characters in paper:

Abbreviations: IS, insertion sequence
Data deposition: The sequences reported in this paper have been deposited in the GenBank database: accession BX936398 (chr.); BX936399 (pYV); BX936400 (pYptb32953).

***Yersinia pestis*, the causative agent of plague, is a highly uniform clone that diverged recently from the enteric pathogen *Yersinia pseudotuberculosis.* Despite their close genetic relationship, they differ radically in their pathogenicity and transmission. Here we report the complete genomic sequence of *Y. pseudotuberculosis* IP32953 and its use for detailed genome comparisons to available *Y. pestis* sequences. Analyses of identified differences across a panel of *Yersinia* isolates from around the world reveals 32 *Y. pestis* chromosomal genes that, together with the two *Y. pestis*-specific plasmids, represent the only new genetic material in *Y. pestis* acquired since the divergence from *Y. pseudotuberculosis.* In contrast, 149 new pseudogenes (doubling the previous estimate) and 317 genes absent from *Y. pestis* were detected, indicating that as many as 13% of *Y. pseudotuberculosis* genes no longer function in *Y. pestis*. Extensive IS-mediated genome rearrangements and reductive evolution through massive gene loss, resulting in elimination and modification of pre-existing gene expression pathways appear to be more important than acquisition of new genes in the evolution of *Y. pestis.* These results provide a sobering example of how a highly virulent epidemic clone can suddenly emerge from a less virulent, closely related progenitor.**

Strong molecular evidence supports the fact that *Y. pseudotuberculosis*, responsible for yersiniosis in animals and humans, is the recent ancestor to *Y. pestis*, the etiologic agent of bubonic and pneumonic plague (1-3). However, while *Y. pseudotuberculosis* is a soil and water-borne enteropathogen, *Y. pestis* is much more dangerous and is of current interest due to its potential use in bioterrorism and as a biological weapon. Present day *Y. pestis* strains, though all similarly pathogenic, can be classified into three biovars (Antiqua, Medievalis, Orientalis) on the basis of their ability to utilize glycerol and to reduce nitrate. These phenotypic differences and molecular typing methods in conjunction with strain geographical origins have served to correlate these biovars with the three recorded plague pandemics.

Of special importance to the pathogenic process of both *Y. pseudotuberculosis* and *Y. pestis* is the shared requirement of a virulence plasmid pCD1 (pYV in enteropathogenic *Yersinia*) that encodes a type III secretion system (4), responsible for injecting into host cells a number of cytotoxins and effectors (Yops) that inhibit bacterial phagocytosis and processes of innate immunity (5, 6). Two additional plasmids unique to *Y. pestis* termed pPCP1 (9.6 kb) and pMT1 (102 kb) play roles in tissue invasion (7, 8) and capsule formation (9) as well as infection of the plague flea vector (10, 11), respectively. However, the presence of these plasmids by themselves cannot account for the remarkable increase in virulence observed in *Y. pestis* (12-15). Despite many extensive studies of the plasmid-encoded virulence determinants induced during the infectious process, and the recent availability of the genome sequences of a *Y. pestis* Orientalis strain, CO92 (16) and a Medievalis strain, KIM10+ (17), the mechanism(s) underlying the strikingly different clinical manifestations of *Y. pseudotuberculosis* and *Y. pestis* have remained elusive. Although a microarray-based comparison of these two Yersinia species has been reported recently (18), the detailed comparison between the completed genomes of *Y. pestis* and that of *Y. pseudotuberculosis* IP32953 (serotype I) presented here provides the first opportunity to examine all differences in genome structure and at the nucleotide level. These comparisons reveal many of the molecular details that were involved in the speciation and emergence of *Y. pestis* and may hold the key to the exceptional virulence of the plague bacillus.

Materials and Methods

Whole genome shotgun libraries were obtained and sequenced as previously described (19). The whole genome sequence of *Y. pseudotuberculosis* IP32953 was obtained from 85,000 end sequences (8.8 fold redundancy), and was assembled using PHRAP (P. Green, University of Washington). All gaps were closed by primer walking on gap-spanning clones or PCR products and a large insert scaffold was used to verify proper genome assembly. Gene modeling and genome annotation was performed as previously described (19). Genome comparisons between the *Yersinia* sequences were viewed using the Artermis Comparison Tool (ACT) (http://www.sanger.ac.uk/Software/ACT/).

The *Yersinia* strains studied came from the collection at the Institut Pasteur. Analysis of these strains was performed by screening PCR results and if necessary, sequencing the resulting products. Specifically, for the strain- or species-specific genes, primers were designed to amplify a ~500 bp region within the gene (or gene portion) that was found to be missing from the other strains or species. For the IS-interrupted genes, primers were designed to amplify a 300-600 bp region of the WT gene or a 1-2.5 kb fragment which includes the interrupting IS element. Due to homologous recombination between IS elements, the alternative and sometimes expected result was a negative one (e.g. no PCR product when the IS in question underwent recombination, or in the event of a deletion removing at least one of the priming sites). For the other pseudogenes, sequencing each PCR product was followed by multiple alignments of the sequences to identify wild-type versus mutant loci. In all cases, experiments yielding negative results were repeated under the same conditions and also using a lower annealing temperature in the event that the region in question had undergone divergence.

The *Y. pestis* and *Y. pseudotuberculosis* genomic DNAs that were used in panel-screens were isolated from the following strains of *Y. pestis* (biovars A - Antiqua, M - Medievalis, O - Orientalis) Harbin (Former Soviet Union - A), Japan (Japan - A), Margaret (Kenya - A), 343 (Belgium Congo - A), PKH-4 (Kurdistan - M), PKR292 (Kurdistan - M), PAR13 (Iran - M), 297RR (Vietnam - O), Exu184 (Brazil - O), Hambourg10 (Germany - O), 6/69 (Madagascar - O); and *Y. pseudotuberculosis*, IP33134 (Russia, serotype I), IP32790 (Italy, I), IP32950 (France, I), IP30215 (Denmark, II), IP32802 (Italy, III), IP32889 (Spain, III), IP31833 (England, IV), IP32952 (France, V). The controls used were *Y. pestis* CO92 (USA - O) and *Y. pseudotuberculosis* IP32953 (France, I). Results for *Y. pestis* KIM10+ were predicted using the available genome sequence.

The sequence of the complete *Y. pseudotuberculosis* strain IP32953 genome is available under the following accession numbers: BX936398 (chromosome); BX936399 (pYV); BX936400 (pYptb32953). Supporting information accompanies this paper and is accessible at www.pnas.org.

Results and Discussion

Genome organization of *Y. pseudotuberculosis* IP32953. The genome of strain IP32953, a fully-virulent clinical isolate from a human patient, consists of a single circular chromosome (4,744,671 bp), the pYV virulence plasmid (68,526 bp) and an atypical novel 27,702 bp cryptic plasmid designated pYptb32953. The general features of the IP32953 genome are listed in Table 1 and the chromosome is represented in Fig. 1. Comparisons of pYV to the previously sequenced pCD1 plasmids from *Y. pestis* KIM5 (20, 21) and *Y. pestis* CO92 (16) revealed an essentially conserved co-linear backbone, differing by the presence in pCD1 of an IS*100* element, a coding sequence (CDS) encoding a 68 amino acid hypothetical protein of unknown function, and an apparent internal in-frame 120 amino acid insertion in the middle of the *yopM* gene, consistent with the known heterogeneity found among YopM in yersiniae (22).

The novel pYptb32953 is likely a conjugative cryptic plasmid and bears similarity to the recently described cryptic plasmid of *Y. enterocolitica* strain 29930 (23). The similarity (~60-65 %) extends to the plasmid mobilization machinery involving TraE, MobB/C homologs, and the entire cluster of type IV conjugation genes involved in plasmid transfer, suggesting that pYptb32953 may be self-transmissible. The latter operon also displays similarity to the conjugation genes of the IncX plasmid R6K (24) and to the *Brucella* spp. *virB* operon (25). Examination of the presence of this plasmid across a large number of isolates belonging to the three pathogenic *Yersinia* species indicated that its distribution is quite narrow (present in three out of 81 strains - Table 2 in supporting information). Thus this plasmid cannot account for any important virulence-associated characteristic of *Y. pseudotuberculosis*.

Few chromosomal features have been known to distinguish *Y. pestis* from *Y. pseudotuberculosis* strains (26, 27). However, comparisons between the chromosome of IP32953 and the *Y. pestis* chromosomes of CO92 and KIM10+ revealed several major differences (Table 1 and Fig.1). The IP32953 chromosome encodes 3,974 predicted genes of which 2,976 (75%) have greater than or equal to 97% identity to their homologues in

Y. pestis. Likewise, the synteny of the *Yersinia* genomes is readily discernable, and the breaks in colinearity have been mapped precisely (Fig.1).

Unique Chromosomal Regions in *Y. pseudotuberculosis*. Thirty-six IP32953-specific regions, ranging in size between 500 bp and 122 kb, are scattered throughout the chromosome and contain a total of 317 putative genes that are not found in either CO92 or KIM10+. A list of putative proteins encoded in these regions and their gene locations can be found in Table 3 (see supporting information). More than one half (188) of the genes in these unique regions are distributed in five clusters composed of phage-like products, the largest of which is a 122 kb region consisting of a 60 kb core of unknown function flanked by two parallel, but different, ~30 kb P2-like mosaic phage regions. Another seven non-phage clusters, encode 49 genes involved primarily in transposition and restriction modification, and together with the phage-associated regions, are also likely to have been horizontally acquired. It appears that the great majority of the remaining clusters encoding 80 genes have been deleted from the *Y. pestis* genome as demonstrated by the presence of partial gene remnants, IS elements, etc. The distribution of these genes (other than phage regions) into functional categories is shown in Fig. 2. Roughly a third of all the IP32953-specific genes in this group are hypothetical or conserved hypothetical genes, while others include genes that encode general metabolic functions that appear to have been lost in *Y. pestis*.

Since species-specific regions and other species-specific genomic features should be conserved across a broad section of strains, a panel of 19 geographically and phenotypically diverse strains of *Y. pestis* and *Y. pseudotuberculosis* was selected and screened for the presence or absence of these features. Of 85 IP32953-specific genes tested by PCR, 11 genes were found to be specific for the *Y. pseudotuberculosis* species (i.e. present in all *Y. pseudotuberculosis* and absent from all *Y. pestis* isolates – Table 4 in supporting information). Only one of these genes, YPTB0537, lies within one of the 12 above-described regions of putative foreign origin. Four of these 11 genes encode hypothetical proteins while four others encode: a putative restriction modification system component (YPTB0537), and proteins involved in glucan biosynthesis (YPTB2493 and YPTB2494) and uracil transport (YPTB2793). The last three encode metabolism-related functions: aspartate aminotransferase, enolase-phosphatase E1, and 5-methylthioribose kinase, respectively. These differences in metabolic enzymes may reflect the differences in *Y. pestis* and *Y. pseudotuberculosis* host ranges. In addition, the *Y. pseudotuberculosis*-specific regions may account for important virulence factors uniquely required in *Y. pseudotuberculosis*, such as the *opg* operon (YPTB2493-YPTB2495) which is required for the synthesis of periplasmic branched glucans that serve in other organisms as an osmoprotectant (28) but that may not be needed in *Y. pestis*, an obligate parasite of eukaryotes, unlikely to experience wide fluctuations in environmental osmotic conditions.

Unique Chromosomal Regions in *Y. pestis*. We also identified 112 KIM10+ and CO92-specific (i.e. not found in IP32953) genes distributed in 21 clusters of 300 bp to 41.7 kb scattered throughout the genome (Table 5 in supporting information). Roughly three categories of genes were identified in these 21 regions: 1) 39 genes (35% of the total) are hypothetical or conserved hypothetical; 2) 59 genes (53%) are phage or transposon-related and 3) 14 genes (12%) can be attributed a putative function. Among those with an ascribed function are membrane proteins, lipoproteins, a putative esterase, a DNA binding protein and a methyltransferase. Our studies indicate that a CO92 nine kb filamentous prophage region, previously believed to be Orientalis biovar specific (18) (27), is in fact also present in some members of the Antiqua biovar [(29) and Table 6 in supporting information] and is absent from IP32953.

Of the 112 genes uniquely associated with the two *Y. pestis* genomes, 105 were tested for their presence or absence in our panel of 19 Yersinia strains (Table 6). Only 32 genes, located in six clusters, were present in all *Y. pestis* and absent from all *Y. pseudotuberculosis* strains examined. Four of these clusters have been recently identified using microarray analysis (18). However, genome sequence comparison coupled with PCR has identified two additional regions not found by hybridization and has eliminated the five other regions previously determined as unique to *Y. pestis* (Table 6) by that method.

Four of the *Y. pestis*-specific gene clusters encode predominantly putative proteins with little, if any, similarity to known or predicted proteins (with the exception of a methylase). Another cluster consists of bacteriophage-related genes (YPO2084-103, YPO2114 in CO92; y2227-y2211, y2201 in KIM10+); while the last cluster (YPO1668-71 in CO92; y1829-y1832) encodes putative membrane proteins, a translation initiation inhibitor, and conserved hypothetical proteins. Though there were no obvious virulence factors encoded in these regions, their role in pathogenicity deserves further study.

Inactivated Genes. Sixty-two pseudogenes are found in IP32953, 43 of which are also pseudogenes in one or both sequenced *Y. pestis* strains (Table 7 in supporting information). The remaining 19 likely represent recent *Y. pseudotuberculosis*-acquired mutations that have arisen since their divergence. Of these, the functions most frequently affected included outer membrane, transport and exported proteins, perhaps reflecting the organism's interaction with its environment. Two of the 19 were integrases with substantial similarity to one another: a P4-like integrase (YPTB0534) and the previously described (30) pathogenicity island HPI integrase (YPTB1602). Though the significance of the other *Y. pseudotuberculosis*-specific inactivated P4-like integrase is not known, the intact counterpart in *Y. pestis*, may be involved in the increased frequency of IS transposition in the latter.

Of the 149 originally reported CO92 pseudogenes (16), only 84 are pseudogenes in the KIM10+ strain and yet are intact genes in IP32953. Three-way gene by gene comparisons among the *Yersinia* strains enabled us to identify 149 additional putative pseudogenes in CO92 (Table 8 in supporting information) of which 124 are also pseudogenes in the KIM10+ genome, yet only 2 are pseudogenes in IP32953. Thus, a closer approximation to the factual number of potentially lost functions by this evolutionary mechanism in *Y. pestis* is 208 (84+124), so that as much as 5% of the gene complement may have been selectively inactivated in *Y. pestis*. A summary of this subset of inactivated genes and their distribution by COG functional classes is shown in Fig. 2.

Using the same panel of 19 strains, we also examined the distribution of 52 randomly-selected CO92 pseudogenes. Forty-six of them could be grouped into five discernable categories, the largest of which comprises 28 pseudogenes specific to *Y. pestis* (Table 9 in supporting information). Members of this group are potentially the most interesting since they affect traits that are unique to *Y. pestis* strains and thus, may represent good targets for studying their novel pathogenic properties and for quick identification in clinical settings. Genes disrupted in this group range from conserved hypothetical, to genes of general metabolism such as *metB* (responsible for the observed methionine requirement of *Y. pestis*), to regulatory genes (e.g. putative two-component sensor kinase, etc.) and potential virulence-associated genes (invasin, toxin transporter, etc.).

A second group of seven pseudogenes was only found in members of the biovar Orientalis, and include the arginine-binding periplasmic protein 2 precursor (*argJ*); the N-terminal region of *E. coli* prepilin peptidase dependent protein (*ppdA*); the exonuclease encoded by *sbcC*; and the aerobic glycerol phosphate dehydrogenase (*glpD*), which is likely responsible for the glycerol-minus phenotype of the biovar Orientalis (31).

Six IS-interrupted pseudogenes comprise a third category, including *aroG*, *pbpC* (penicillin-binding protein 1C) and *setA*, a sugar efflux transporter. These are pseudogenes in all members of the Orientalis biovar as well as in one or both of the African Antiqua strains (from Kenya and Congo), and are intact genes in *Y. pseudotuberculosis*, the Medievalis lineage, and the non-African Antiqua strains. This finding, in addition to the previously alluded to filamentous phage distribution pattern supports the notion that the Orientalis and Medievalis lineages arose independently from Antiqua biovar.

A fourth category of two other pseudogenes, a putative surface protein (YPO0902 in CO92, y3288 in KIM10+) and a pectin degradation protein (YPO1726 in CO92, y1888 in KIM10+) are found in all *Y. pestis* strains and are also present in several *Y. pseudotuberculosis* strains. These may represent mutations acquired prior to the emergence of *Y. pestis*, as they are unlikely to have been independently acquired by each species (one is a partial deletion and the other is interrupted by an IS*285*).

The fifth and last category comprises a single IS*100*-interrupted acetylornithine aminotransferase, *argD*, a CO92-specific pseudogene, likely the result of a very recent insertion sequence mobility that supports the idea of a continuously fluid genome.

Metabolism. Since *Y. pseudotuberculosis* is a chemoheterotroph, a full complement of biosynthetic and intermediary metabolic pathways was expected and has been verified. As already indicated, several of the IP32953-specific regions encode general metabolic functions and thus, may account for some of the observed physiological differences between the two species. Noteworthy among this group are genes of purine and aspartate metabolism as well as of the methionine salvage pathway (32-34). Gene inactivations that may account for the *Y. pestis*-specific biochemical phenotypes include: a cysteine synthase (*cysM*) frameshift (the cysteine requirement of *Y. pestis*), a missense point mutation affecting amino acid 363 in the aspartate ammonia-lyase (*aspA*) of *Y. pestis* likely accounting for the stimulatory effect of CO_2 on growth (35), and a proline substitution present in amino acid 161 of glucose 6 P-dehydrogenase (*zwf*) that likely prevents utilization of hexose via the pentose-phosphate pathway (36). The significance of these last two types of mutations will require further functional analyses.

Pathogenicity. Genomic differences that may play a role in the unique pathogenic characteristics of these two species include alterations in lipid A biosynthesis exemplified by the absence in *Y. pestis* of lipid A acyltransferase gene *htrB* (YPTB2490), which adds an acyl group to lipid A, and may account for the differences in lipid A between the two species. Since lipid A acylation changes are known to alter endotoxic properties and interactions with the innate immune system, this difference could be of significance for pathogenesis.

Several hemolysins/hemagglutinins homologues of different pathogens are present in the yersiniae. In IP32953, a cluster of nine CDS (YPTB3450- YPTB3459) encode several hemolysin homologues in a region absent from *Y. pestis*. A hemolysin activator is a pseudogene in both IP32953 (YPTB3651) and KIM10+ (y0002) but is wild-type in CO92 (YPO3720). However, since this mutation occurs at a homopolymeric tract of C's (11 in IP32953 and KIM10+, and only 7 in CO92), it may simply represent a spontaneous reversion, similar to that shown to occur in *ureD* in which silencing and reactivation of urease in *Y. pestis* is determined by a spontaneous addition/excision of a single G residue in the *ureD* gene (37). Another hemolysin gene that is inactivated by partial deletion in IP32953 (YPTB2524) and all other *Y. pseudotuberculosis* strains is found intact in *Y. pestis* (Table 6; gene YPO2486 in CO92 and y1701 in KIM10+). Although the role of hemolysins in *Yersinia* virulence remains unclear, their conserved nature and clear differences among the species suggest the need for further studies to investigate their possible function.

The insecticidal toxin homologs found in either complete or inactivated form in the *Y. pestis* genomes have been implicated in the adaptation of this organism to the flea life cycle (16, 18, 38). Thus, it has been suggested that the observed inactivation of *tcaB*, encoding an insecticidal toxin protein, is required for flea life cycle but this argument can now be refuted as this gene is complete and normal in several Medievalis and Antiqua *Y. pestis* strains (Table 9). Similarly, the in-frame deletion of *tcaC* in *Y. pestis* cannot alone account for its ability to colonize the flea midgut, as this gene is even shorter in *Y. pseudotuberculosis*, neither can the same function be attributed to the viral enhancing protein previously described in CO92 (16), since it is also present in IP32953. Thus, the precise role of insecticidal toxin homologs in flea midgut colonization remains largely unresolved.

Two loci (*srfA* and *srfB*) encoding putative virulence factors, along with the gene for the Cu-Zn superoxide dismutase, *sodC* have in-frame insertion/deletions in KIM10+ and CO92, but are wild-type in IP32953. If these mutations affect protein function, they could play a role in species-specific virulence. Similarly, an IS*1541* neighboring *csrB,* a small non-coding RNA that antagonizes CsrA, an S-layer protein involved in adherence to cells, may modify the transcription and/or stability of this RNA and thus may have an effect on virulence in *Y. pestis.* Another region that could have a role in virulence in these organisms is a high pathogenicity island-like region, HPI-2 (noted in the CO92 genome, accession AL590842). This region is wild-type in *Y. pseudotuberculosis* but defective in *Y. pestis* in which the siderophore synthesis protein (YPO0778 and YPO1012 in CO92; y3406 and y3410 in KIM10+) is inactivated by an IS*100* insertion.

Regulatory Genes. At least nine regulatory genes that are inactivated in *Y. pestis* could have effects on its phenotype, including virulence. A frameshift in *Y. pestis*-homologues of YPTB0553 affects a gene similar to *sorC*, a transcriptional regulator required for sorbose utilization, while a frameshift in the *Y. pestis* homologues of YPTB1259 may affect the regulation of the synthesis of polysaccharide colanic acid. This capsular polysaccharide has been implicated in blocking the specific binding between uropathogenic *E. coli* and inert substrates (39). These inactivations in *Y. pestis* may be consistent with the general loss of adhesins that are unnecessary for its life-style. The gene *flhD* may be one of many genes inactivated in *Y. pestis* responsible for altered motility in this organism. Its absence may have a positive impact on Yop expression (40) and a possible pleiotropic effect on virulence and metabolism as demonstrated in other enterobacteria (41). Also inactivated in *Y. pestis* is the *rhafR* homologue (YPO1728 in CO92; y2579 in KIM10+) which may lead to derepression of the rhaffinose utilization pathway in this organism.

A frameshift in the transcriptional regulator, *iclR* carried by *Y. pestis* leads to constitutive glyoxylate bypass in this organism explaining an already known phenomenon (42). Furthermore, since the glyoxylate bypass has been shown to be necessary for virulence in other bacterial pathogens and fungi (43, 44), constitutive expression may also enhance *Y. pestis* virulence.

UhpB (YPTB3846), a transcriptional activator of genes involved in the uptake and metabolism of hexose phosphates, is inactivated in many *Y. pestis* strains. Finally, the gene encoding sigma N modulating factor (YPTB3527) possesses a stop codon in position 36 in *Y. pestis*, which could lead to modified expression of sigma 54 dependent genes.

IS Elements, Genome Rearrangements and Evolution. Only 20 IS elements were found in the IP32953 chromosome, in stark contrast to the 117 in KIM10+ and 138 in CO92 (Table 1). Twelve of the 20 IS elements in IP32953 share integration locations with those in the two *Y. pestis* strains, suggesting that only 8 recent transposition events have occurred in IP32953, whereas an extraordinary expansion of each IS family took place in *Y. pestis* strains since their divergence. Examination of the shared IS locations within CO92 and KIM10+ suggests that their most recent common ancestor carried 109 IS elements and that since the divergence of this ancestral representative and the present day KIM10+ and CO92 strains, 8 and 28 new insertions occurred respectively. What remains unclear is whether the rate of transposition in *Yersinia* is periodically stimulated or if these events occurred in a punctuated fashion upon some as yet unknown induction.

Despite the dense distribution of IS elements in *Y. pestis* and their potential for generating homologous recombination-mediated deletions, there are surprisingly few (only five) IP32953-specific regions that can be the result of excision of intervening sequence via recombination at flanking direct IS elements in *Y. pestis*.

Deng *et al.* (17) first alluded to the important role played by repeat elements (namely IS elements) in explaining the unique genome arrangement displayed by the two sequenced *Y. pestis* strains. Analyses using the structural organization of IP32953 for comparison further support the role played by IS elements in genome evolution and confirms the ancestral character of *Y. pseudotuberculosis*, since IP32953 most often has no "equivalent" IS element when compared with *Y. pestis*. This implies that most rearrangements have occurred only recently in the *Y. pestis* lineage and that the genome structural organization of IP32953 more closely reflects that of the ancestral type.

In a manner analogous to that utilized in the KIM10+/CO92 comparisons (17), we can identify some 32 syntenic colinear blocks conserved between IP32953 and CO92 (Fig. 1) and 25 between IP32953 and KIM10+. The genome organization of the last common ancestral genome of the two *Y. pestis* strains, as well as the ancestral genome of both species, could be deduced by investigating the precise locations of these rearrangements. Thus, IP32953 has undergone at least one and likely no more than three intra-chromosomal recombinations since the split from the last common ancestor. A probable recombination in IP32953 that generated a large inversion between two IS*1661* is supported by the distinct shift in GC skew associated with this region (Fig. 1). Two other putative IP32953 rearrangements are exemplified by the mobile pathogenicity island HPI, which typically integrates at one of three *asn*-tRNAs in *Yersinia* spp. (45), and a recombination at a P4-like integrase (YPTB0534) that is common to all three sequenced strains. All other rearrangements appeared to have occurred in the *Y. pestis* lineage.

Allowing for the three possible rearrangements proposed during the evolution of IP32953, the progenitor of both CO92 and KIM10+, must have undergone at least 11 recombination/rearrangement events (undoubtedly influenced by the 97 additional IS elements gained since diverging from *Y. pseudotuberculosis*). KIM10+ and CO92 have since undergone an additional 10 and 18 rearrangements respectively, commensurate to their respective increased levels of IS transposition. It is thus quite likely that the insertion elements themselves and/or the subsequent rearrangements they have generated have played an important role in the emergence of *Y. pestis* from its *Y. pseudotuberculosis* ancestor.

Implications in virulence and pathogen evolution

The genome sequence of *Y. pseudotuberculosis* IP32953 and its comparison to *Y. pestis* reveal aspects of the evolutionary processes that evidently transformed a common enteropathogenic ancestor, and later gave rise to two present-day pathogens of vastly distinct clinical manifestations. Molecular events that likely operated during the evolution of alternatively free-living *Y. pseudotuberculosis* (capable of causing localized chronic disease) contrast with those involved in the evolution of *Y. pestis* (capable of causing vector-dependent acute disease). The extensive chromosomal rearrangements that occurred during the emergence of *Y. pestis* undoubtedly are indicative of the mechanisms that drove the evolution of this pathogen. IS element expansion and its corollary, the increased fluidity of the genome, together with massive gene inactivation almost surely have played a role in this process. A direct comparison between the work presented in this study and the calculated evolutionary distances between these two *Yersinia* species presented by Achtman *et al*. (1) is difficult to make without a reliable molecular clock to measure the rates of genome rearrangement and IS transposition and gene inactivation. Since the mechanism(s) that account for IS element expansion and increased gene inactivation in *Y. pestis* is unknown, we can only surmise that these processes were driven by selection for lethality as well as evolutionary pressures that further enabled colonization of the flea. In this scenario, gene inactivation or IS-mediated rearrangements (either before or after the lateral transfer of pPCP1 and pMT1) might have led to changes that increased virulence (high septicemia) and facilitated flea-borne transmission. The concomitant and dramatic change in lifestyle undergone by *Y. pestis*, ensuing from its continuous association with the host and dependency on the flea vector for survival, would have been sufficient to provide the selective pressure that resulted in wholesale inactivation of as much as 13% of its genome that we observe today. This may represent an intermediate stage in genome compaction, a process that has been proposed in the evolution of other pathogens closely associated with their hosts such as *Salmonella typhi* (46) and *Mycobacterium lepae* (47). Finally, the significance of horizontal gene transfer into the chromosome of *Y. pestis* is uncertain. It may be hypothesized that the acquisition of at least some of the six chromosomal regions uniquely conserved in *Y. pestis* strains, in conjunction with the high degree of gene inactivation has been responsible for the increased pathogenicity of this species. Whole genome comparisons of pathogen nearneighbors of distinct characteristics, such as those described in this study, lay the foundation for future mutational, functional and animal studies that will ultimately help elucidate the mechanisms underlying the emergence of new pathogens.

We thank S. Falkow, D. Monack, P. Agron, M. Chu and C. Kim for helpful discussions and review of the manuscript. This work was performed under the auspices of the U.S. DOE by UC, LLNL under Contract No. W-7405-ENG-48 and ORNL under Contract DE-AC05-00OR22725, and by contract 99 01 110 00 470 94 50 from the French Délégation Générale à l'Armement.

1. Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A. & Carniel, E. (1999) *Proc Natl Acad Sci U S A* **96,** 14043-8.
2. Brenner, D. J., Steigerwalt, A. G., Falcao, D. P., Weaver, R. E. & Fanning, G. R. (1976) *Int. J. Syst. Bacteriol* **26,** 180-194.
3. Moore, R. L. & Brubaker, R. R. (1975) *Int. J. Syst. Bact.* **25,** 336-339.
4. Cornelis, G. R. & Van Gijsegem, F. (2000) *Annu Rev Microbiol* **54,** 735-74.
5. Brubaker, R. R. (2003) *Infect Immun* **71,** 3673-81.
6. Cornelis, G. R. (2002) *Nat Rev Mol Cell Biol* **3,** 742-52.
7. Brubaker, R. R., Beesley, E. D. & Surgalla, M. J. (1965) *Science* **149,** 422-424.
8. Lahteenmaki, K., Virkola, R., Saren, A., Emody, L. & Korhonen, T. K. (1998) *Infect Immun* **66,** 5755-62.
9. Kutyrev, V. V., Popov Iu, A. & Protsenko, O. A. (1986) *Mol Gen Mikrobiol Virusol***,** 3-11.
10. Hinnebusch, B. J., Rudolph, A. E., Cherepanov, P., Dixon, J. E., Schwan, T. G. & Forsberg, A. (2002) *Science* **296,** 733-5.
11. Hinnebusch, B. J. (2003) *Adv Exp Med Biol* **529,** 55-62.
12. Filippov, A. A., Solodovnikov, N. S., Kookleva, L. M. & Protsenko, O. A. (1990) *FEMS Microbiol Lett* **55,** 45-8.
13. Friedlander, A. M., Welkos, S. L., Worsham, P. L., Andrews, G. P., Heath, D. G., Anderson, G. W., Jr., Pitt, M. L., Estep, J. & Davis, K. (1995) *Clin Infect Dis* **21 Suppl 2,** S178-81.
14. Kutyrev, V., Mehigh, R. J., Motin, V. L., Pokrovskaya, M. S., Smirnov, G. B. & Brubaker, R. R. (1999) *Infect Immun* **67,** 1359-67.
15. Welkos, S. L., Andrews, G. P., Lindler, L. E., Snellings, N. J. & Strachan, S. D. (2004) *Plasmid* **51,** 1-11.
16. Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C., Mungall, K. L., *et al.* (2001) *Nature* **413,** 523-7.
17. Deng, W., Burland, V., Plunkett, G., 3rd, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., *et al.* (2002) *J Bacteriol* **184,** 4601-11.
18. Hinchliffe, S. J., Isherwood, K. E., Stabler, R. A., Prentice, M. B., Rakin, A., Nichols, R. A., Oyston, P. C., Hinds, J., Titball, R. W. & Wren, B. W. (2003) *Genome Res* **13,** 2018-29.
19. Chain, P., Lamerdin, J., Larimer, F., Regala, W., Lao, V., Land, M., Hauser, L., Hooper, A., Klotz, M., Norton, J., *et al.* (2003) *J Bacteriol* **185,** 2759-73.
20. Hu, P., Elliott, J., McCready, P., Skowronski, E., Garnes, J., Kobayashi, A., Brubaker, R. R. & Garcia, E. (1998) *J Bacteriol* **180,** 5192-202.
21. Perry, R. D., Straley, S. C., Fetherston, J. D., Rose, D. J., Gregor, J. & Blattner, F. R. (1998) *Infect Immun* **66,** 4611-23.
22. Boland, A., Havaux, S. & Cornelis, G. R. (1998) *Microb Pathog* **25,** 343-8.
23. Strauch, E., Goelz, G., Knabner, D., Konietzny, A., Lanka, E. & Appel, B. (2003) *Microbiology* **149,** 2829-45.
24. Nunez, B., Avila, P. & de la Cruz, F. (1997) *Mol Microbiol* **24,** 1157-68.

25. Boschiroli, M. L., Ouahrani-Bettache, S., Foulongne, V., Michaux-Charachon, S., Bourg, G., Allardet-Servent, A., Cazevieille, C., Lavigne, J. P., Liautard, J. P., Ramuz, M. & O'Callaghan, D. (2002) *Vet Microbiol* **90,** 341-8.
26. Brubaker, R. R. (2000) in *The prokaryotes, an evolving electronic resource for the microbiological community, vol. 2000.*, ed. Stackelbrandt, E. (Springer Verlag, New York.).
27. Radnedge, L., Agron, P. G., Worsham, P. L. & Andersen, G. L. (2002) *Microbiology* **148,** 1687-98.
28. Kennedy, E. P. (1996) in *Escherichia coli & Salmonella*, ed. Neidhardt, F. C. (ASM Press, Washington DC), Vol. I, pp. 1064-1071.
29. Gonzalez, M. D., Lichtensteiger, C. A., Caughlan, R. & Vimr, E. R. (2002) *J Bacteriol* **184,** 6050-5.
30. Lesic, B., Bach, S., Ghigo, J. M., Dobrindt, U., Hacker, J. & Carniel, E. (2004) *Mol Microbiol* **52,** 1337-48.
31. Motin, V. L., Georgescu, A. M., Elliott, J. M., Hu, P., Worsham, P. L., Ott, L. L., Slezak, T. R., Sokhansanj, B. A., Regala, W. M., Brubaker, R. R. & Garcia, E. (2002) *J Bacteriol* **184,** 1019-27.
32. Mortlock, R. P. (1962) *J Bacteriol* **84,** 53-9.
33. Brubaker, R. R. (1970) *Infect Immun* **1,** 446-454.
34. Dreyfus, L. A. & Brubaker, R. R. (1978) *J Bacteriol* **136,** 757-64.
35. Baugh, C. L., Lanham, J. W. & Surgalla, M. J. (1964) *J Bacteriol* **88,** 553-8.
36. Mortlock, R. P. & Brubaker, R. R. (1962) *J Bacteriol* **84,** 1122-1123.
37. Sebbane, F., Devalckenaere, A., Foulon, J., Carniel, E. & Simonet, M. (2001) *Infect Immun* **69,** 170-6.
38. Wren, B. W. (2003) *Nat Rev Microbiol* **1,** 55-64.
39. Hanna, A., Berg, M., Stout, V. & Razatos, A. (2003) *Appl Environ Microbiol* **69,** 4474-81.
40. Bleves, S., Marenne, M. N., Detry, G. & Cornelis, G. R. (2002) *J Bacteriol* **184,** 3214-23.
41. Pruss, B. M., Campbell, J. W., Van Dyk, T. K., Zhu, C., Kogan, Y. & Matsumura, P. (2003) *J Bacteriol* **185,** 534-43.
42. Hillier, S. & Charnetzky, W. T. (1981) *J Bacteriol* **145,** 452-8.
43. Lorenz, M. C. & Fink, G. R. (2002) *Eukaryot Cell* **1,** 657-62.
44. Lorenz, M. C. & Fink, G. R. (2001) *Nature* **412,** 83-6.
45. Buchrieser, C., Brosch, R., Bach, S., Guiyoule, A. & Carniel, E. (1998) *Mol Microbiol* **30,** 965-78.
46. Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D., Wain, J., Churcher, C., Mungall, K. L., Bentley, S. D., Holden, M. T., *et al*. (2001) *Nature* **413,** 848-52.
47. Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., *et al*. (2001) *Nature* **409,** 1007-11.

Fig. 1. Circular genome map of IP32953 and comparison with *Y. pestis* CO92.

Panel A, Genome of IP32953; Panel B, Genome of CO92. For both panels, circle 1 (from center outward), G+C content; circles 2 and 3, all genes coded by function (forward and reverse strand); circle 4, GC skew (G-C/G+C); circles 5 and 6, genome divided into locally colinear blocks (when IP32953 and CO92 are compared with one another), each block is distinguished by a unique color (black segments within colored blocks represent regions specific to that genome in the comparison) and the orientation of each block is indicated by strand (circle 5, -ve strand; 6, +ve strand); circle 7, locations of IS elements (blue IS*100*, red IS*285*, green IS*1661*, magenta IS*1541*). In panel A, the gray highlighted region near 12 o'clock indicates the proposed IP32953 inversion (see text) while the remainder of the genome denotes the stable "ancestral" arrangement that has prevailed through the present. Panel B illustrates the complexity of the molecular events that gave rise to the inversions or translocations in the *Y. pestis* genome first proposed (16) solely on the basis of the dramatic shifts in G/C skew (gray highlights I, II and III) but now extended through whole genome comparison. For example, gray highlight II is composed of three distinct blocks, two derived from distinct places within the same replichore (origin to terminus half) while the third one originated from the other replichore (light blue block).

Fig. 2. Functional classification of genes missing or inactivated in *Y. pestis*.

Distribution of *Y. pestis*-specific lost functions by gene region deletion (light blue) or by gene inactivation (i.e. pseudogene, dark purple) in COG functional groups: C, Energy production; D, Cell division, chromosome partitioning; E, Amino acid metabolism; F, Nucleotide metabolism; G, Carbohydrate metabolism; H, Coenzyme metabolism; I, Lipid metabolism; J, Translation; K, Transcription; L, DNA replication, repair; M, Cell envelope biogenesis; N, Cell motility, secretion; O, Posttranslational modification; P, Inorganic ion metabolism; R, General function prediction only; S, Function unknown; T, Signal transduction; conserved, conserved hypothetical genes with no significant COG hits; unique, hypothetical genes with no significant COG hit.

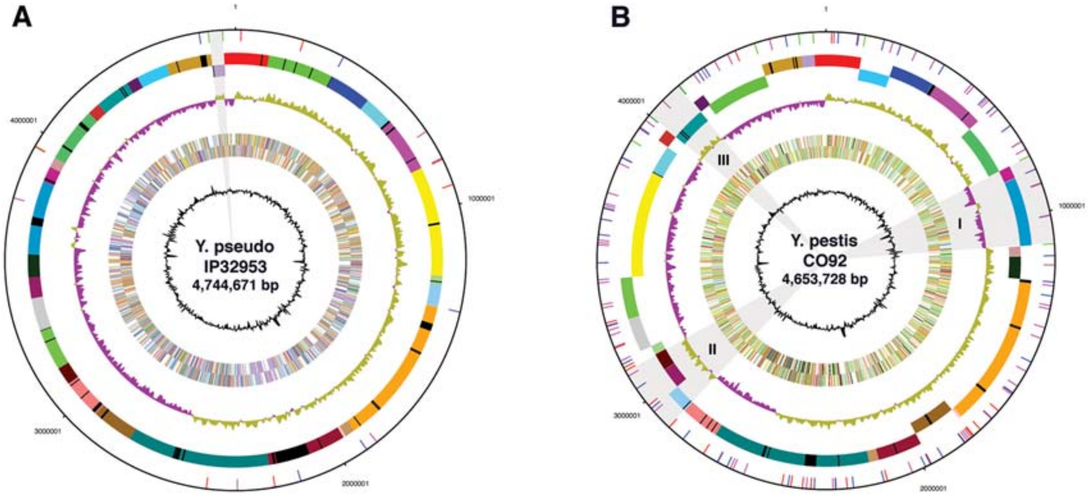
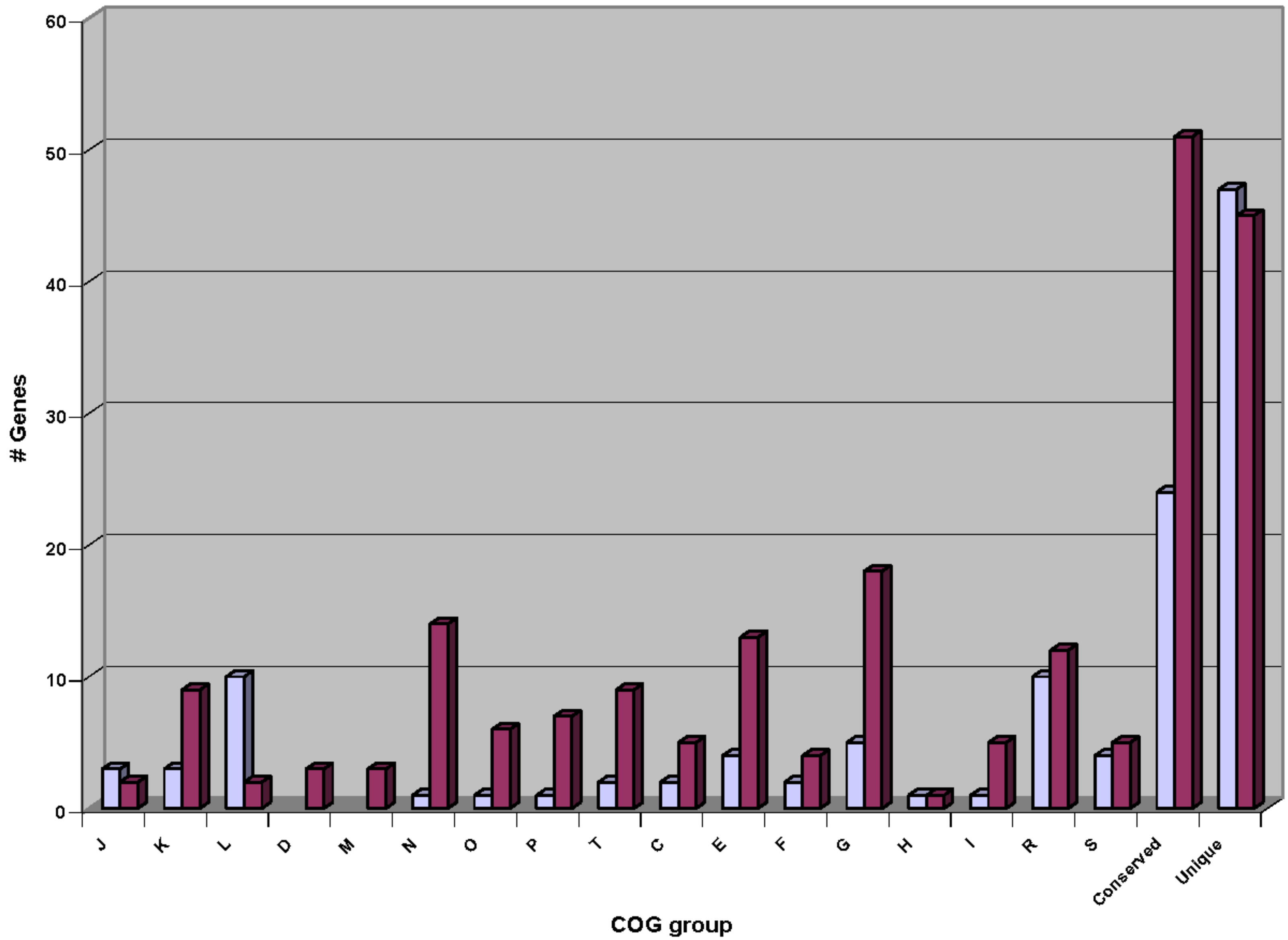

A
B
1
1000001
2000001
3000001
4000001
Y. pseudo
IP32953
4,744,671 bp
Y. pestis
CO92
4,653,728 bp
I
II
III

Fig. 2



Genes
60
50
40
30
20
10
0
J
K
L
D
M
N
O
P
T
C
E
F
G
H
I
R
S
Conserved
Unique
COG group

Table 2: Distribution of the 27 kb plasmid among pathogenic *Yersinia*

Y. pseudotuberculosis (50 strains)	Origin	Country	Biotype	Serotype	P126	P127	P128*	P129*	P130*
IP32953	Human	France	NA	I	+	+	+	+	+
IP32790	Pig	Italia	NA	I	-	-			
IP32954	Human	France	NA	I	-	-			
IP32637	Unknown	France	NA	I	-	-			
IP32950	Human	France	NA	I	-	-			
IP32929	Hare	France	NA	II	-	-			
IP32934	Monkey	France	NA	II	-	-			
IP32951	Human	France	NA	II	-	-			
IP32921	Hare	France	NA	II	-	-			
IP32937	Bovine	Argentina	NA	III	-	-			
IP32938	Bovine	Argentina	NA	III	-	-			
IP32509	Unknown	Spain	NA	IV	-	-			
IP31411	Hare	Denmark	NA	IV	-	-			
IP32816	Hare	Japan	NA	V	-	-			
IP32817	Hare	Japan	NA	V	-	-			
IP32821	Human	France	NA	V	-	-			
IP31553	Guinea-pig	Japan	NA	VI	-	-			
IP32949	Human	France	NA	I	-	-			
IP32889	Unknown	Spain	NA	III	+	-	-	+	+
IP31833	Sheep	England	NA	IV	-	-			
IP32952	Human	France	NA	V	-	-			
IP30151	Otter	Sweden	NA	IV	-	-			
IP30215	Guinea pig	Danemark	NA	II	-	-			
IP30284	Pigeon	Italy	NA	I	-	-			
IP30437	Beaver	Canada	NA	I	-	-			
IP30642	Mouse	Tunisia	NA	I	-	-			
IP30911	Hare	Holland	NA	II	-	-			
IP31554	Guinea pig	Japan	NA	VI	-	-			
IP31878	Rodent	Tunisia	NA	I	-	-			
IP32323	Water	Norway	NA	II	-	-			
IP32463	Guinea pig	Switzerland	NA	V	-	-			
IP32533	Deer	New Zealand	NA	I	-	-			
IP32581	Human	Belgium	NA	II	-	-			
IP32665	Hare	Yougoslavia	NA	I	-	-			
IP32802	Pig	Italy	NA	III	-	-			
IP32939	Soil	Romania	NA	I	-	-			
IP32984	Human	Spain	NA	III	-	-			
IP32992	Bovine	Australia	NA	III	-	-			
IP33005	Monkey	Germany	NA	I	+	+	+	+	+
IP33012	Monkey	Germany	NA	II	-	-			
IP33023	Monkey	Switzerland	NA	II	-	-			
IP33038	Marsupial	Australia	NA	I	-	-			
IP33051	Caprine	France	NA	III	-	-			
IP33054	Human	Sapin	NA	II	-	-			
IP33061	Monkey	Germany	NA	V	-	-			
IP33088	Human	France	NA	II	-	-			
IP33097	Deer	Argentina	NA	III	-	-			
IP33098	Hare	France	NA	II	-	-			
IP33105	Bovine	Argentina	NA	III	-	-			
IP33108	Human	Bulgaria	NA	III	-	-			
IP33109	Human	France	NA	I	-	-			

Y. pestis (15 strains)	Origin	Country	Biotype	Serotype	P126	P127	P128*	P129*	P130*
564	Rodent	Kurdistan	M	NA	-	-			
613	Unknown	Burma	O	NA	-	-			
554	Human	Kenya	A	NA	-	-			
544	Human	Kenya	A	NA	-	-			
549	Human	Belgium Congo	A	NA	-	-			
520	Rodent	Kurdistan	M	NA	-	-			
665	Human	Madagascar	O	NA	-	-			
1273	Human	Vietnam	O	NA	-	-			
1222	Human	Vietnam	O	NA	-	-			
507	Unknown	Vietnam	O	NA	-	-			
1053	Human	Vietnam	O	NA	-	-			
663	Human	Madagascar	O	NA	-	-			
249	Human	Madagascar	O	NA	-	-			
570	Rodent	Brazil	O	NA	-	-			
501	Unknown	Germany	O	NA	-	-			
***Y. enterocolitica* (15 strains)**	**Origin**	**Country**	**Biotype**	**Serotype**	**P126**	**P127**	**P128***	**P129***	**P130***
IP21591	Food	Australia	1	O:5	-	-			
IP21642	Human	France	1	O:5	-	-			
IP66	Chinchilla	Switzerland	3	O:1,2a,3	-	-			
IP69	Chinchilla	Switzerland	3	O:1,2a,3	-	-			
IP3	Hare	France	5	O:2a,2b,3	-	-			
IP21461	Human	France	2	O:9	-	-			
IP21412	Food	Argentina	2	O:9	-	-			
IP22274	Human	Australia	4	O:3	-	-			
IP21699	Human	France	4	O:3	-	-			
IP845	Human	U.S.A.	1B	O:20	-	-			
IP21605	Food	Australia	1A	O:7,13	-	-			
IP21440	Food	France	1A	O:7,8,19	-	-			
IP21389	Food	France	1A	O:7,8,13,19	-	-			
IP21373	Food	France	1A	O:6,31	-	-			
IP21633	Food	France	1A	O:10,34	-	-			

*Primers were only tested in the 3 *Y. pseudotuberculosis* strains
NA: Not applicable

Primer Sequences

P126	F	ATAAGCTGGCTGGCAAATATG
	R	GAAGGTCAGACGGTGACTGAG
P127	F	AATAATCACCTCTAGGAGG
	R	TCTTCTTTGTCACTAGGGTT
P128	F	ATGTTAGCGCCAGAATTGGA
	R	GTACGCCTTTATCACGTGCC
P129	F	ATGAGTGCTATTCTCTCGGA
	R	ACATGCGCCTCGCCTATTAG
P130	F	CGCCAAGAAAATCCCCTTGG
	R	ATGAGGTATCTGTAACTTGG

Table 3: Regions and genes (or domains of functional genes) specific to IP32953, in comparisons with CO92 and KIM10+

Regions	Gene	Location	Product	Possible IS-mediated deletions in *Y. pestis*	Putative regions of lateral transfer
1	YPTB0149	181028..181282	putative colicin immunity protein		phage
	YPTB0150	181435..181854	putative pyocin S2		
	YPTB0151	181856..182110	pyocin S2 immunity protein		
	YPTB0152	182263..182691	putative pyocin S2		
	YPTB0153	182688..182951	putative colicin immunity protein		
	YPTB0154	183509..183880	probable phage antitermination protein Q		
	YPTB0155	184491..184826	hypothetical		
2	YPTB0187	complement(227114..228151)	possible Pentapeptide repeats protei		
3	YPTB0244	287857..288525	hypothetical		
	YPTB0245	288538..289740	conserved hypothetical protei		
4	YPTB0535	complement(628794..632042)	putative type I restriction enzyme, R subunit		other
	YPTB0536	complement(632052..633335)	putative type I restriction enzyme, S subunit		
	YPTB0537	complement(633337..635928)	putative type I restriction-modification system, methyltransferase subunit (N-6 DNA Methylas		
5	YPTB0557	658664..659932	possible conserved cysteine desulfurase		
	YPTB0558	659979..661157	possible acyl-CoA dehydrogenase		
	YPTB0559	661159..661434	hypothetical protein		
	YPTB0560	661436..662272	hypothetical protein		
	YPTB0561	662305..663531	putative protein involved in molybdopterin biosynthesis		
	YPTB0562	complement(663713..663949)	putative transposase		
6	YPTB0664	794132..795634	hypothetical protein		other
	YPTB0665	complement(795879..796010)	conserved hypothetical protein		
	YPTB0666	complement(796037..796414)	putative IS1400 transposase B		
7	YPTB0872	complement(1047598..1048368)	putative amidase-type enzyme		
	YPTB0873	complement(1048356..1049516)	putative aspartate aminotransferase		
	YPTB0874	1049656..1050324	probable sugar aldolase		
	YPTB0875	1050321..1051010	enolase-phosphatase E-1		
	YPTB0876	1051109..1051651	methionine salvage pathway enzyme E-2/E-2'		
	YPTB0877	complement(1051684..1052724)	putative translation initiation factor EIF-2B, GDP-GTP exchange factor (alpha subunit)		
	YPTB0878	1052920..1054143	5-methylthioribose kinase		
8	YPTB1058	1265226..1266098	conserved hypothetical protei	IS*285*	
9	YPTB1202	1432494..1433750	xanthosine permease	IS*1541*	
	YPTB1203	complement(1433838..1434311)	conserved hypothetical protein		
	YPTB1204	1434661..1436475	possible 2-component Histidine kinase-sensor		
	YPTB1205	1436462..1437826	transcriptional regulatory protein hydG		
	YPTB1206	1438174..1439283	morphinone reductase		
	YPTB1207	complement(1439284..1440195)	putative transcriptional regulator		
	YPTB1208	1440765..1441091	conserved hypothetical protein		
	YPTB1209	complement(1441130..1442236)	conserved hypothetical protein		
	YPTB1210	complement(1442249..1443424)	conserved hypothetical protein		
	YPTB1211	complement(1443429..1445171)	putative ATP-binding component of a transport system		
	YPTB1212	complement(1445185..1446171)	putative membrane protein		
	YPTB1213	complement(1446223..1446933)	putative transcriptional regulator		
	YPTB1214	1447324..1448679	putative ATP-dependent RNA helicase rhlE		
10	YPTB1287	1535703..1536419	bacteriophage tail fiber protein		phage
	YPTB1288	1536421..1536840	putative tail fiber assembly protein p3		
11	YPTB1292	complement(1540821..1541105)	putative colicin immunity protein		other
	YPTB1293	complement(1541134..1541391)	conserved hypothetical protei		
12	YPTB1491	1787812..1789083	hypothetical		
	YPTB1492	1789090..1789482	putative lipoprotein		
	YPTB1493	1789798..1792500	conserved hypothetical protein		
	YPTB1494	1792503..1792919	hypothetical		
	YPTB1495	1792928..1794676	possible LysM domain		
	YPTB1496	1794681..1795193	hypothetical		
13	YPTB1564	complement(1882973..1883629)	hypothetical		
14	YPTB1690	complement(2046009..2046608)	hypothetical		
	YPTB1738	complement(2097207..2097422)	prophage p2 ogr protein		
	YPTB1739	complement(2097514..2098680)	similar to D protein bacteriophage 186 and P2		
	YPTB1740	complement(2098677..2099162)	putative phage tail protein		
	YPTB1741	complement(2099162..2101588)	putative bacteriophage P2 tail protein gpT		
	YPTB1742	complement(2101581..2101703)	gpE+E' [Enterobacteria phage P2] gb|AAD03292.1| (AF063097) g...		
	YPTB1743	complement(2101736..2102047)	putative tail protein gpE P2 bacteriophage		
	YPTB1744	complement(2102098..2102613)	putative tail tube protein FII (bacteriophage P2)		
	YPTB1745	complement(2102627..2103796)	putative tail sheath protein (P2 and 186 bacteriophage)		
	YPTB1746	complement(2103924..2104406)	conserved hypothetical phage tail fiber protein		
	YPTB1747	complement(2104418..2105857)	hypothetical phage tail fiber protein		
	YPTB1748	complement(2105854..2106462)	putative bacteriophage protein		
	YPTB1749	complement(2106455..2107363)	probable bacteriophage protein		
	YPTB1750	complement(2107368..2107718)	possible phage-related protein		
	YPTB1751	complement(2107715..2108356)	putative phage-related baseplate assembly protein		
	YPTB1752	complement(2108430..2108879)	O protein [Enterobacteria phage 186] gb|AAC34159.1| (U32222)...		
	YPTB1753	complement(2108876..2109331)	gpR [Enterobacteria phage P2] sp|P36933|VPR_BPP2 TAIL COMPLE...		
	YPTB1754	complement(2109427..2109843)	putative Orf27; P2 LysB homolog; control of lysis [Ente...		
	YPTB1755	complement(2109845..2110351)	putative phage lysozyme		
	YPTB1756	complement(2110335..2110556)	likely phage related protein		
	YPTB1757	complement(2110547..2110750)	putative phage-related tail protein		
	YPTB1758	complement(2110750..2111223)	gpL [Enterobacteria phage P2] sp|P25475|VPL_BPP2 HEAD COMPLE...		
	YPTB1759	complement(2111323..2111982)	R protein [Enterobacteria phage 186] gb|AAC34151.1| (U32222)...		
	YPTB1760	complement(2111986..2113224)	(AJ298563) major capsid protein [Bacteriophage PhiD266]		
	YPTB1761	complement(2113301..2114155)	similar to V protein bacteriophage 186		
	YPTB1762	2114304..2116076	terminase subunit [Enterobacteria phage 186] gb|AAC3414...		
	YPTB1763	2116073..2116837	putative W protein [Enterobacteria phage 186] gb|AAC34147.1|...		
	YPTB1764	2116834..2117871	capsid portal protein [Enterobacteria phage 186] gb|AAC...		
	YPTB1765	complement(2118623..2118982)	putative gp46 [Bacteriophage N15] pir||T13133 protein gp46 -...		
	YPTB1766	complement(2119005..2121287)	putative phage P2 replication protein		
	YPTB1767	complement(2121274..2121546)	hypothetical protein 79 phage 186		
	YPTB1768	complement(2121612..2121923)	similar to gpB bacteriophage P2		
	YPTB1769	complement(2122130..2122639)	regulatory protein CII bacteriophage 186		
	YPTB1770	complement(2122694..2122891)	putative regulator for prophage CP-933T (E. coli O157:H7)		
	YPTB1771	2122962..2123573	CI repressor of phage 186 and others		
	YPTB1772	2123601..2124872	hypothetical protein		
	YPTB1773	2124872..2125750	hypothetical		
	YPTB1774	2125870..2126925	putative integrase		
	YPTB1775	2126922..2127968	hypothetical		
	YPTB1776	complement(2128303..2129325)	putative integrase		
	YPTB1777	complement(2129645..2130148)	hypothetical		
	YPTB1778	complement(2130242..2130580)	hypothetical		
	YPTB1779	complement(2130712..2131575)	hypothetical		
	YPTB1780	complement(2131934..2132170)	hypothetical		
	YPTB1781	complement(2132407..2132835)	hypothetical		
	YPTB1782	complement(2132832..2133971)	hypothetical		
	YPTB1783	complement(2134016..2135620)	conserved hypothetical protein		
	YPTB1784	complement(2135607..2136449)	conserved hypothetical protein		
	YPTB1785	complement(2136750..2137088)	hypothetical protein		
	YPTB1786	complement(2137092..2137598)	hypothetical protein		
	YPTB1787	complement(2138463..2138732)	hypothetical protein		
	YPTB1788	complement(2138817..2139062)	hypothetical protein		
	YPTB1789	complement(2139340..2139957)	hypothetical protein		

15	YPTB1790	complement(2139968..2140654)	putative prophage repressor protein		phage
	YPTB1791	2140734..2141024	hypothetical protein		
	YPTB1792	2141042..2141404	hypothetical protein		
	YPTB1793	2141589..2142434	putative phage protein		
	YPTB1794	2142438..2143214	hypothetical protein		
	YPTB1795	2143218..2144354	possible recombination associated protein RdgC		
	YPTB1796	2144354..2144701	hypothetical protein		
	YPTB1797	2144698..2145624	putative DNA methyltransferase		
	YPTB1798	2145621..2146496	conserved hypothetical protein		
	YPTB1799	2146493..2148238	putative modification methylase		
	YPTB1800	2148235..2148882	conserved putative phage protein		
	YPTB1801	2149276..2150211	hypothetical protein		
	YPTB1802	2150208..2150666	hypothetical protein		
	YPTB1803	2150978..2151787	hypothetical protein		
	YPTB1804	2151960..2152172	putative holin protein		
	YPTB1805	2152172..2152660	probable endolysin (lysis protein) (lysozyme)		
	YPTB1806	2152792..2153310	hypothetical protein		
	YPTB1807	2153528..2153968	putative UNKNOWN PROTEIN [Lactococcus lactis subsp. lac...		
	YPTB1808	2154075..2155097	putative transposase IS100		
	YPTB1809	2155097..2155876	putative IS100 transposase		
	YPTB1810	2155994..2156851	putative small subunit bacteriophage terminase		
	YPTB1811	2156848..2158554	putative bacteriophage terminase large subunit		
	YPTB1812	2158554..2160674	putative phage portal protein		
	YPTB1813	2160944..2161993	putative phage l protein		
	YPTB1814	2162053..2163276	putative phage protein		
	YPTB1815	2163354..2163737	putative phage protein		
	YPTB1816	2163803..2164255	putative phage protein		
	YPTB1817	2164258..2164854	conserved hypothetical protein		
	YPTB1818	2164854..2165507	putative phage protein		
	YPTB1819	2165504..2167639	hypothetical phage protein		
	YPTB1820	2167650..2168114	putative phage tail fiber assembly protein		
	YPTB1821	complement(2168127..2168453)	putative acyl carrier protein		
	YPTB1822	complement(2168447..2168911)	putative membrane protein		
	YPTB1823	2169143..2170771	phage hypothetical protein		
	YPTB1824	2170768..2171874	bacteriophage hypothetical protein		
	YPTB1825	2171941..2172972	conserved hypothetical protein		
	YPTB1826	2172947..2173321	bacteriophage hypothetical protein		
	YPTB1827	2173329..2173934	bacteriophage hypothetical protein		
	YPTB1828	2173950..2174333	bacteriophage hypothetical protein		
	YPTB1829	2174333..2174572	bacteriophage hypothetical protein		
	YPTB1830	2174583..2175974	bacteriophage hypothetical protein		
	YPTB1831	2176308..2182709	bacteriophage hypothetical protein		
	YPTB1832	2182663..2184330	hypothetical protein		
	YPTB1833	complement(2184581..2184985)	hypothetical		
	YPTB1834	complement(2185864..2186037)	prophage p2 ogr protein		
	YPTB1835	complement(2186112..2187206)	bacteriophage P2 gpD protein		
	YPTB1836	complement(2187203..2187667)	putative bacteriophage P2 tail protein		
	YPTB1837	complement(2187679..2190594)	putative tail fiber component T of bacteriophage P2		
	YPTB1838	complement(2190587..2190709)	gpE+E' [Enterobacteria phage P2] gb|AAD03292.1| (AF063097) g...		
	YPTB1839	complement(2190721..2191068)	possible (AF153829) unknown [Salmonella typhi]		
	YPTB1840	complement(2191122..2191637)	putative P2 tail tube protein		
	YPTB1841	complement(2191649..2192824)	putative bacteriophage P2 tail sheath protein		
	YPTB1842	complement(2192958..2193437)	conserved hypothetical protein		
	YPTB1843	complement(2193449..2194888)	bacteriophage hypothetical protein		
	YPTB1844	complement(2194885..2195493)	putative bacteriophage P2 protein		
	YPTB1845	complement(2195486..2196394)	putative bacteriophage P2 related protein		
	YPTB1846	complement(2196397..2196750)	putative phage-related protein		
	YPTB1847	complement(2196747..2197382)	Orf32; P2 V homolog; baseplate protein [Enterobacteria ...		
	YPTB1848	complement(2197713..2198429)	hypothetical protein		
	YPTB1849	complement(2198746..2199192)	putative O protein [bacterio phage 186] gb|AAC34159.1|...		
	YPTB1850	complement(2199189..2199656)	gpR [Enterobacteria phage P2] sp|P36933|VPR_BPP2 TAIL COMPLE...		
	YPTB1851	complement(2199755..2200180)	putative (AB008550) orf12; similar to LysB gene of P2:l...		
	YPTB1852	complement(2200185..2200580)	bacteriophage P7 related protein		
	YPTB1853	complement(2200567..2200953)	conserved hypothetical protein		
	YPTB1854	complement(2200983..2201186)	putative WPhiphage-related tail protein		
	YPTB1855	complement(2201186..2201677)	putative orf4 of phage P2 (gene L)		
	YPTB1856	complement(2201911..2202564)	putative R protein bacteriophage 186 gb|AAC34151.1|		
	YPTB1857	complement(2202571..2203626)	putative major capsid protein [Bacteriophage PhiD5]		
	YPTB1858	complement(2203663..2204478)	similar to V protein phage 186		
	YPTB1859	2204649..2206412	similar to gpP phage P2 TERMINASE		
	YPTB1860	2206460..2207452	similar to capsid portal protein bacteriophage 186 gb|AAC...		
	YPTB1861	complement(2208134..2208478)	putative Bacteriophage protein gp46		
	YPTB1862	complement(2208686..2211241)	putative phage replication protein		
	YPTB1863	complement(2211330..2212148)	putative DNA adenine methylase		
	YPTB1864	complement(2212145..2212798)	similar to Orf81 bacteriophage 186		
	YPTB1865	complement(2212855..2213622)	hypothetical protein		
	YPTB1866	complement(2213705..2214013)	hypothetical protein		
	YPTB1867	complement(2214041..2214301)	hypothetical protein		
	YPTB1868	complement(2214443..2214757)	hypothetical protein		
	YPTB1869	complement(2214972..2215175)	putative DNA-binding protein		
	YPTB1870	2215440..2215808	putative prophage transcriptional regulator		
	YPTB1871	2215829..2217526	similar to hypothetical bacteriophage P27 protein		
	YPTB1872	2217547..2217966	putative bacteriophage integrase		
	YPTB1873	2218035..2218415	hypothetical protein		
	YPTB1874	2218415..2218747	putative bacteriophage integrase		
	YPTB1875	2219041..2219418	putative prophage integrase		
16	YPTB1884	2225288..2226646	possible multidrug-efflux transporter		other
	YPTB1885	2226643..2228376	possible ThiF family		
	YPTB1886	2228366..2229091	conserved hypothetical protein		
	YPTB1887	2229079..2229933	hypothetical protein		
	YPTB1888	2229918..2230730	conserved hypothetical protein		
	YPTB1889	2230727..2231893	possible diaminopimelate decarboxylase		
	YPTB1890	2231868..2232425	putative similar to ribosomal-protein-alanine N-acetyltransfe...		
	YPTB1891	complement(2232455..2232757)	weak similarity to Vibrio transposas		
17	YPTB2180	complement(2565873..2566877)	adenosine deaminase		
	YPTB2181	2567266..2568522	hypothetical		
	YPTB2182	complement(2568695..2569624)	conserved hypothetical protein		
	YPTB2183	2569923..2570747	putative transcriptional regulator		
18	YPTB2193	complement(2583135..2583380)	conserved hypothetical protein		
	YPTB2194	complement(2583594..2583968)	conserved hypothetical protein		
	YPTB2195	complement(2584195..2584944)	probable oxidoreductase in dcp-nohA intergenic region		
	YPTB2196	complement(2585058..2586860)	putative transcriptional regulator		
	YPTB2197	2587156..2588676	probable aldehyde dehydrogenase		
	YPTB2198	complement(2588957..2590555)	(AF335466) unknown [Yersinia pseudotuberculosis]		
	YPTB2199	2591079..2592002	putative membrane protein		
19	YPTB2200	complement(2592808..2593986)	putative aminotransferase		
	YPTB2201	complement(2593973..2594542)	conserved hypothetical protein		
20	YPTB2205	complement(2597263..2598324)	ribose ABC transporter, permease protein		
	YPTB2206	complement(2598351..2599889)	putative ribose ABC transporter, ATP-binding protein		
	YPTB2207	complement(2599886..2600539)	conserved hypothetical protein		
	YPTB2455	2900644..2900910	putative phage replication protein gene fragment		
	YPTB2456	2901382..2901567	hypothetical		

21	YPTB2457	complement(2901687..2902217)	hypothetical	IS*1541*	phage
	YPTB2458	complement(2902330..2903445)	hypothetical		
	YPTB2459	complement(2903918..2905096)	hypothetical		
	YPTB2460	complement(2905610..2906083)	putative capsid portal protein [Enterobacteria phage 18...		
	YPTB2461	2906112..2906390	hypothetical		
	YPTB2462	2906975..2907220	conserved hypothetical protein		
22	YPTB2490	2936394..2937314	lipid A biosynthesis lauroyl acyltransferase		
	YPTB2491	2938011..2939549	putative di-tripeptide transport system permease protein		
	YPTB2492	complement(2939796..2940128)	conserved hypothetical protein		
	YPTB2493	complement(2940161..2942770)	membrane glycosyltransferase		
	YPTB2494	complement(2942763..2944349)	periplasmic glucans biosynthesis protein		
	YPTB2495	2944676..2945572	glucans biosynthesis protein (partial)		
	YPTB2496	2945778..2946023	hypothetical		
	YPTB2497	2946367..2948094	conserved hypothetical protein		
23	YPTB2539	complement(3006939..3007601)	conserved hypothetical protein		
24	YPTB2749	complement(3249394..3250119)	hypothetical		other
	YPTB2750	complement(3250122..3250829)	hypothetical		
	YPTB2751	complement(3250826..3251497)	hypothetical		
	YPTB2752	3251579..3251869	part of IS630		
	YPTB2752	3251854..3251988	part of IS630 orfA - pseudogene		
	YPTB2752	3252037..3252162	part of IS630 orfA - pseudogene		
25	YPTB2793	complement(3295818..3297104)	uracil transport	IS*1541*	
26	YPTB3129	complement(3688397..3689380)	Possible phage integrase/recombinase		phage
	YPTB3130	complement(3689447..3689794)	Putative [Enterobacteria phage P2] gpC-like protein.		
	YPTB3131	3689860..3690132	Probable bacteriophage Cox protein		
	YPTB3132	3690338..3690622	hypothetical protein		
	YPTB3133	3691181..3691450	Hypothetical		
	YPTB3134	3691680..3692030	Conserved hypothetical protein		
	YPTB3135	3692337..3693002	putative conserved bacteriophage protein		
	YPTB3136	3693012..3693713	putative conserved bacteriophage protein		
	YPTB3137	3694060..3696615	Putative phage protein.		
	YPTB3138	complement(3696877..3697203)	Possible (AB008550) orf34 [bacteriophage phi CTX]		
	YPTB3139	complement(3697200..3698267)	Hypothetical, similar to orf34 (AB008550) [phage phi CTX].		
	YPTB3140	complement(3698264..3700069)	Possible [Haemophilus phage HP1] orf16-like phage protein.		
	YPTB3141	3700242..3701330	hypothetical phage protein		
	YPTB3142	3701365..3702393	probable phage protein		
	YPTB3143	3702396..3703100	Conserved hypothetical phage protein		
	YPTB3144	3703209..3703682	Conserved hypothetical phage protein		
	YPTB3145	3703679..3704173	Conserved hypothetical phage protein		
	YPTB3146	3704170..3704829	Conserved hypothetical phage protein		
	YPTB3147	3704853..3706019	Conserved hypothetical phage protein		
	YPTB3148	3706022..3706477	Conserved hypothetical protein		
	YPTB3149	3706487..3706789	possible phage P2 holin-like protein		
	YPTB3150	3706786..3707127	Hypothetical		
	YPTB3151	3707127..3707501	hypothetical		
	YPTB3152	3707616..3707891	hypothetical protein similar to protein 26 - phage HP1		
	YPTB3153	3708117..3710093	Putative bacteriophage tail protein.		
	YPTB3154	3710090..3710419	hypothetical protein similar to protein 28 - phage HP1		
	YPTB3155	3710416..3711600	Putative bacteriophage protein.		
	YPTB3156	3711593..3712216	Putative bacteriophage protein.		
	YPTB3157	3712226..3713821	Hypothetical protein		
	YPTB3352	3983362..3983775	Flagellar switch protein		
	YPTB3159	3714336..3715046	Possible bacteriophage protein.		
	YPTB3160	3715039..3715593	Hypothetical protein similar to protein 34 (AY27935) - Phage HP2		
	YPTB3161	3715590..3717311	Hypothetical bacteriophage protein		
27	YPTB3274	complement(3859378..3859902)	hypothetical		
	YPTB3275	complement(3860127..3860465)	hypothetical		
	YPTB3276	complement(3860595..3861122)	hypothetical		
28	YPTB3279	complement(3864435..3865094)	Hypothetical		
	YPTB3280	complement(3865483..3865908)	Hypothetical		
	YPTB3281	complement(3866252..3866743)	Hypothetical		
29	YPTB3364	complement(3997470..3998000)	fimbriae	IS*1541*	other
	YPTB3365	3999120..4000535	Putative Serratia marcescens-like C1 chitinase.		
	YPTB3366	4000604..4002013	chitin-binding protein		
	YPTB3367	4002384..4002578	Possible transposase fragment.		
	YPTB3368	4003499..4004086	Possible Yersinia enterocolitica-like Orf1 (AF005744)		
	YPTB3369	4004309..4004506	Hypothetical		
30	YPTB3451	complement(4098135..4098386)	Hypothetical		
	YPTB3452	complement(4098388..4098705)	conserved hypothetical protein		
	YPTB3453	complement(4098990..4099343)	hypothetical		
31	YPTB3455	complement(4101394..4101894)	conserved hypothetical protein		
32	YPTB3458	complement(4108387..4108776)	hypothetical		
	YPTB3459	complement(4108782..4109594)	hypothetical protein		
33	YPTB3620	complement(4306508..4307125)	hypothetical protein		
34	YPTB3613	complement(4297082..4297327)	rhs accessory		
	YPTB3614	complement(4297516..4297797)	hypothetical		
35	YPTB3795	complement(4526145..4527395)	conserved hypothetical protein		
36	YPTB3862	complement(4615019..4615396)	putative DNA-binding protein		other
	YPTB3863	complement(4615368..4615775)	conserved hypothetical protein		
	YPTB3865	complement(4616645..4617502)	conserved hypothetical protein		
	YPTB3866	complement(4617726..4618190)	conserved hypothetical protein		
	YPTB3867	complement(4618147..4618482)	conserved hypothetical protein		
	YPTB3868	complement(4618494..4618973)	conserved hypothetical protein		
	YPTB3869	complement(4619156..4619776)	hypothetical		
	YPTB3870	complement(4619804..4620358)	hypothetical		
	YPTB3871	complement(4620606..4621238)	conserved hypothetical protein		
	YPTB3872	complement(4621286..4621615)	putative IS100 transposase		
	YPTB3873	4622244..4622864	conserved hypothetical protein		
	YPTB3874	complement(4623046..4623696)	mrr restriction system protein		
	YPTB3877	complement(4625594..4625920)	mrr restriction system protein		
	YPTB3878	complement(4626056..4626793)	putative predicted metal-dependent hydrolase		
	YPTB3879	complement(4626797..4630075)	possible type I restriction enzyme (restriction subunit)		
	YPTB3880	complement(4630072..4631118)	conserved hypothetical protein		
	YPTB3881	complement(4631115..4632344)	possible restriction modification enzyme		
	YPTB3882	complement(4632334..4633857)	putative type I site-specific deoxyribonuclease LldI chain hs...		
	YPTB3883	complement(4633864..4634442)	hypothetical		
	YPTB3884	4634660..4635574	conserved hypothetical protein		
	YPTB3885	complement(4635877..4636101)	putative (AJ250469) putative SinR-like protein [Clostri.		

Table 4: Identification of *Y. pseudotuberculosis*-specific genes across a panel of strains of the two species

Regions from ST2*	IP32953 genes	*Y. pestis strains* (O, Orientalis; M, Medievalis; A, Antiqua)													*Y. pseudotuberculosis strains* (Serovars I through V)									Oligonucleotides used for PCR	
		O	O	O	O	O	M	M	M	M	A	A	A	A	I	I	I	I	II	V	III	III	IV		
		CO92	297RR	Hambourg10	Exu184	6/69	KIM10+	PKH-4	PKR292	PAR13	Harbin	Japan	Margaret	343	IP32953	IP33134	IP32790	IP32950	IP30215	IP32952	IP32802	IP32889	IP31833	Primer 1	Primer 2
1-phage	YPTB0155	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	+w	+	-	+	TGCAGTGGAGTAACCTGTCAA	TGTTGAGCAGCACAAAAATCA
	YPTB0150	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	+w	+w	-	+	GGCGTTTTGGAAGGTAGTTG	TCTATATGCCTTTTCGGAGTT
	YPTB0153	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	+	-	+	ACGGAAGAAGAATTTATTGATTT	CTTTTTCAAGTATCTTCCTTGGTG
2	YPTB0187	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	+w	-	+	CCGGAGCTGAGTTAGATTTCA	CGATATGCCATTACGGGATT
3	YPTB0244	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+w	+	-	+w	+	+	+	AGTTCGGCTGGTGTCAAAAT	TGCCGTTTTAGTGTAGGCAA
	YPTB0245	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+	-	+w	+	+	+	CTAACCCAAATGCGAGAAGC	ACGCTACCTTGTTCAAACCG
4-other	YPTB0536	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	TATTTCCCCCGAGGTTTACA	TTGTGTAATGCGCAATCGTT
	YPTB0535	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	+	+w	+	AAGGGCTGGGAAACAGAGAT	GCTGCGCTTCAGCTCTAAAT
	YPTB0537	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+*	+	+	+	+	+	+	GCCAACCAGATGCGTTCTAA	TCAGGTGCAGTAAGTCGCTG
5	YPTB0558	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	-	-	-	GGACGGGTGCAACTGATATT	GCACCAAATTTACTTCCCGA
	YPTB0561	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	+	-	-	GCCCCGTTTTGCTTTATTTA	CCAGCCTCAGCACAATTAGG
6-other	YPTB0664	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+	+	+	+	-	-	TGGCACCATTGAAGATGAAA	ACGTTACCGCATGTTTCTCC
7	YPTB0873	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	CAGCGGCTAGCCTATCATGT	TCACTTCATCACCAGGACGA
	YPTB0875	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	TGACACCTTTCTTACGTGCG	ATAAGCGCTCACGTCTCGTT
	YPTB0877	-	-	-	+w	-	-	-	-	-	-	-	+w	-	+	+	+	+	+	+	+	+	+	GCTACTGGCTGAACGTGGTT	CGACCCAGATAGCATCCACT
	YPTB0878	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	CCGGGTGATTGTTAAACAGG	GATCTTCGGTGATTTGGCAT
8	YPTB1058	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	ACAGCAAGCGTGTCTTTCCT	GGGCTTCTTTCTGCACTTTG
9	YPTB1202	-	-	-	-	-	-	-	+w	+w	+	+	+	+	+	+	+	+	+	+	+	+	+	ACGCTGCGCTTCACTCTATT	CTATTCCAGTCCTTGACGCC
	YPTB1206	+	-	-	+	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CTATTCGTGCGGGTTTTGAT	GCAATATAAGGACGCCCAAA
	YPTB1210	+w	-	+w	+	+w	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	GGCGTTTATCGAATTGGTTC	AGTAAAATGCCCACATGCAA
	YPTB1214	+w	-	+w	+	+w	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	CCTCAATTCATGCCTCGTCT	CACCACGCAATTTCATCATC
10-phage	YPTB1287	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+w	+	-	+w	+	+	+	TGGGTGTGGCAACTAATTCTC	CCAAGTTCTCCCATGGTGTT
	YPTB1288	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+	-	+	+	+	+	CCCGATGGAAAAACACTAGG	TTTATTGGCCATGTGATTGG
12	YPTB1491	-	-	-	+w	-	-	-	-	-	-	-	+w	-	+	+	+	-	-	-	+	-	-	GGATGTTTATGGAAAACGAATTT	TCACGGGAGGAATTATCACC
	YPTB1495	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+w	+	+	+w	-	+w	GAAGCCGATGATGGTTCAAT	CAAGCGTCACTTTCAGTGGA
13	YPTB1564	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	+	-	+	GGCAAAAGCTATTCGCTACG	TCAGCCATCCCTCCAATAAG
14	YPTB1690	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+	+	+	+w	+	+	GGCTTGGCAATGGTGTTTAT	TTCCCTTGGGTAAAAGACCC
15-phage	YPTB1739	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	CCCAGAATATTCGTGATCGG	TTGCGTCTGGTCGATATGAG
	YPTB1745	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	ACATTTCCCCTTAACACCCC	CGGTACAACATCGCTTCTGA
	YPTB1751	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	TGGCCTGTATATCGCCTTTC	CCTTGGGTTACATTCAGCGT
	YPTB1760	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	AGTGGGGTTGAGTGTTACCG	CAAGTGCCTGTCGCTTGATA
	YPTB1764	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	GCTCACCGCTGTATGTGAAA	GTTGCGAAAATTGCCTAACC
	YPTB1772	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	CAGGGCCTAATGGTTATGGA	GCTTCAAATGCTTCAACGAA
	YPTB1775	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	CCCTCCGGTATGGAGCTATT	TTCCTGAGCAGACTCCCAAT
	YPTB1782	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	ACCTTGCCATTGAGCAGACT	CACGCGTTGATCAGAAATTG
	YPTB1795	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	CCTGTGGTCCCTTTGACACT	TCGTCCCGTTCTACTCCATC
	YPTB1800	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+w	-	-	-	-	-	-	-	GTATTGGGCGTGCTGTTACC	TGACGTGGTTGAAATAAGCG
	YPTB1806	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+w	-	-	-	-	-	-	AAAAGCGAGGAAAGGCAAAT	GCTTGGCGGTATGTCATTCT
	YPTB1812	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	CCCTTGACGCAACATAACCT	ATTTGCTCGGCATAAGCATC
	YPTB1831	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	GAAATCTCAAGATCCGAGCG	GCCACGCATCTTTATCGAAT
	YPTB1837	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	-	+	+	CATCCATCAAAACGACGAAA	TGTGGAGTGCGGTACGTAAG
	YPTB1843	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+w	+	+	+w	+	+	GCTGCCCTGAATTCGTTAAG	GTGACTGCCTCGGTGCTACT
	YPTB1859	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	+w	+	+	TGGATAGCTGGGAAAAATGG	GCCGCCATGAATGATATTTT
	YPTB1871	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	+w	-	-	+	CAATGCGGAATTATTTGCAG	TACGCCGAACCTGCTATTTC
16-other	YPTB1885	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+w	-	-	+w	-	-	CAGGACAGACTTTCCATGACC	GTCGCACGGGTAAGGTTAGA
	YPTB1887	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	CGATTCCCTTGCGAAAACTA	CGGGAATGGTTCTTTGGATA
	YPTB1889	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	CTCAGTTCGTTCGCGAATTT	CGATAGTCGGAACTTAGGCG
17	YPTB2181	-	-	-	-	-	-	-	-	-	-	-	+w	-	+	+	+	+	+w	+w	+	+	+	ACCAACCCATTCCAAGTCAA	GACGTCGAAGCTCCATGTTT
	YPTB2182	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	+w	+	+	+	ATCCTGAGTTTGCTGTCGCT	TCCTCTTCGCTGACAATGTG
18	YPTB2193	-	-	-	+w	-	-	-	-	-	-	-	+w	-	+	+	+	+	+	+	+	+	+	TGGCATCTTGCGAAAGTGTA	ATCGGTGATGCGGGTATAAA
	YPTB2194	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	CCTGCTGATGTTGCCATTAGT	CCAGGCAACGCTGAGTATTA
	YPTB2196	-	-	-	-	-	-	-	-	-	-	-	+w	-	+	+	+	+	-	+w	+w	+	+	ACGGTTTAGATCGCATGTCC	GGGCAAACTGTTGTGCTTTT
	YPTB2198	-	-	-	+w	-	-	-	-	-	-	-	+w	-	+	+	+	+	-	+w	+	+	-	ACGCTATTTTCGGCATGAAT	GCAACAGTCCCGCTAATTGT
	YPTB2199	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	GTCATTGGCAGGGGATTCTA	AGCTCTCCAAACTTACGGCA
19	YPTB2201	-	-	-	+w	-	-	-	+w	-	-	-	+w	-	+	+	+	+	+	+	+	+	+	TCAGTGACGTTCCTCGTCTG	CATGGGATCAAAGGCTCAAT
20	YPTB2205	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	-	+	+	GCTTTCGCGCTTATTACTGG	TACGCCAAGTGACTCAATGC
	YPTB2207	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+w	+	+	+	+	+	+	AGCAATGTGTCTGAACTTTCTTCA	CATATTGCCGTCACCGACTACACC
21-phage	YPTB2455	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	CTGGAAGATGAAGACGGGAA	ACTGTTGTGCATGGAATGGA
	YPTB2458	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	GGCTGTAAGTGGCTTGGAGT	TCAACCTGACCATTGGAACC
	YPTB2460	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	CCGGGGTGATTGTTTATCTG	AGGGCTTTCAGACCTTCCAT
	YPTB2462	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	GCATCCATTAAGTCCGGATG	CAACCTGTCCGTAAGCCATT
22	YPTB2491	-	-	-	+w	-	-	-	-	-	-	-	+w	-	+	+	+	+	+	+	+	+	+	GAGCGTTTCTCGTTCTACGG	CGGCTTATACAAGGTCCCAA
	YPTB2493	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	ACAGGACACCAAGCACACAG	CGCGTAACCTATCCCAAAAA
	YPTB2494	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	ATTCTCCACCTCTCAGGCAC	ATAGCTGGCACCGAGCATAC
	YPTB2497	-	-	-	+w	-	-	-	-	-	-	-	+w	-	+	+	+	+	+	+	+	+	+	CGGAATCGCGTTTGATAACT	CGGAATCGCGTTTGATAACT
24-other	YPTB2749	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	+	-	+	CAGTAGATGCAGCGGAGAAA	GCTCAGCCCAGTAACCATCT
	YPTB2751	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	+	-	+	GCTACAGCGTTAGGGCTTTG	TGCCTCGAGTCAATTCTCCT
25	YPTB2793	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	TCTGTAAGGGCAAAATTCCG	CCAATGAGCATCACGAATTG
26-phage	YPTB3129	-	-	-	-	-	-	-	-	-	-	-	-	-	+w	-	-	-	-	-	-	+w	-	TTTAGCGGTTCTTAGCGGAA	ACCCTGGGGTAGGTCTGTCT
	YPTB3135	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+	-	GGGATGATTTGATTTTTGCG	CCCGCTATCCCATTCATTTA
	YPTB3140	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	CGGCGCGAACACTTTATATT	TGGCCCAACTATCCAACATT
	YPTB3145	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	ATAACGCGGAACTGATTCCA	AACTCCACGACGATTTCGAG

	YPTB3153	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+	-	GGCCGCATTTAAACAGATTG	CCATCGCAGACATATAGCCA
	YPTB3157	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+	-	CCGTCCGGACACTGTTATTT	TTATCACCCTCCCTGACGAG
	YPTB3161	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+	-	GTCGGTGATGATGTGATTGC	AACCACTGCCATTGTGTTGA
27	YPTB3274	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	+w	-	+	+	CTTGGTTGCTACCAGAACGG	CACAAAAGACTGACGCCCTT
	YPTB3276	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	+	+	CTTTGGTTGATGGTTCTTGGA	CCCAATGAGTGTCCTTTTCC
29-other	YPTB3365	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+w	-	+w	+w	-	+	CCGGCAAATGGCTATAAAGA	TAAGTGAGGGAATTCGGGTG
	YPTB3368	-	-	-	+w	-	-	-	-	-	-	-	+w	-	+	+	+	+	+	+	-	+	+	GTGGCGGTATTAACCGAAAA	GCAGGATCGTGCATGATAGA
31	YPTB3455	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	GTATGGTCGCTGCGGATAAT	CCCAACCTTCCAGTTTTTCA
35	YPTB3795	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+w	-	-	+w	+	-	+	CAAGTTTGGTCGAAGTGGTG	CCTTCAGCCTGGAAAGGTACT
36-other	YPTB3865	-	-	-	+w	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	-	-	-	GGCGACCTACCAGCTATCAA	GATTGTCAGGAACGGGCTTA
	YPTB3871	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	-	-	-	TATCGGCAGTTGGTTTCACA	GTCACCTTCTCCGTCCGATA
	YPTB3878	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	-	-	-	GCCACTATTACCTGGGCAAA	TGGGATTCAGAGTCACCCTC
	YPTB3880	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	-	-	-	CGAGCACCTGATCAACATCT	GCTTGTTCTGGATGGTTTGAA
	YPTB3885	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	-	-	-	TCTGCGGAAGCAGAAAGATT	ACCGTCTCTCCGTAAACCAGT

Identified in this study as *Y. pestis*-specific genes

+ : amplification of expected size

+w : weak amplification of expected size

- : no amplification

* Regions possibly horizontally-acquired (phage or other origin) are indicated

Table 5: Regions and genes (or domains of functional genes) specific to CO92 and KIM10+ when compared with IP32953

	CO92 annotation			KIM10+ annotation			Putative regions of lateral transfer
Regions	Gene	Location	Product	Gene	Location	Product	
1	YPO0387	403049..405112	hypothetical protein	y3797	complement(4215478..4217544)	hypothetical	
	YPO0388	405109..406425	conserved hypothetical protein	y3796	complement(4214165..4215481)	hypothetical	
	YPO0389	406492..406722	hypothetical protein	NC			
	YPO0390	406969..407436	hypothetical protein	y3794	complement(4213154..4213621)	hypothetical	
	YPO0391	407449..408774	modification methylase	y3792	complement(4211816..4213141)	putative methyltransferase	
	YPO0392	408824..409546	hypothetical protein	y3791	complement(4211044..4211766)	hypothetical	
	YPO0393	409656..410891	hypothetical protein	y3790	complement(4209699..4211072)	hypothetical	
	YPO0394	complement(411235..411522)	hypothetical protein	y3789	4209068..4209355	hypothetical	
	YPO0396	412139..413551	hypothetical protein	y3786	complement(4207039..4208427)	hypothetical	
	YPO0397	413548..414702	hypothetical protein	y3785	complement(4205888..4207042)	hypothetical	
2	YPO0523	567541..567825	putative membrane protein	NC			
3	YPO1087	1234752..1235063	putative prophage protein	NC			phage
	YPO1088	1235071..1235391	putative DNA-binding prophage protein	y3089	complement(3399434..3399754)	hypothetical	
	YPO1089	complement(1235740..1236828)	putative regulatory prophage protein	y3088	3397991..3399085	hypothetical	
	YPO1090	complement(1236825..1237784)	putative prophage DNA primase	y3087	3397011..3398000	hypothetical	
	YPO1091	complement(1237777..1238319)	putative prophage protein	y3086	3396506..3397048	hypothetical	
	YPO1092	complement(1238512..1239408)	putative DNA-binding prophage protein	y3085	3395417..3396313	hypothetical	
	YPO1092a	complement(1239421..1239711)	hypothetical protein	y3084	3395114..3395404	hypothetical	
	YPO1094	complement(1239803..1239976)	hypothetical protein	y3083	3392560..3394050	hypothetical	
	YPO1095	1240207..1240434	hypothetical protein	NC			
	YPO1096	complement(1240596..1242266)	putative phage protein	y3083	3392560..3394050	hypothetical	
	YPO1097	complement(1242317..1243405)	putative phage protein	y3082	3391421..3392509	hypothetical	
	YPO1098	complement(1243484..1244683)	putative prophage integrase	y3081	3390110..3391342	prophage CP4-57 integrase	
4	YPO1252	1409403..1410182	putative bacteriophage tail fiber protein	y2934	complement(3237524..3238303)	hypothetical	phage
5	YPO1257	complement(1414135..1414623)	hypothetical protein	y2929	3233074..3233571	hypothetical	
	YPO1258	complement(1414636..1415784)	conserved hypothetical protein	y2928	3231922..3233070	hypothetical	
6	YPO1473	1668162..1670462	hypothetical protein	y2697	complement(2977173..2979473)	hypothetical	other
	YPO1474	1670459..1671232	putative exported protein	y2696	complement(2976403..2977176)	hypothetical	
	YPO1475	1671386..1671646	hypothetical protein	y2695	complement(2975989..2976249)	hypothetical	
	YPO1476	1671662..1673845	hypothetical protein	y2694	complement(2973790..2975952)	hypothetical	
	YPO1477	1674018..1674488	hypothetical protein	y2692	complement(2973147..2973617)	hypothetical	
	YPO1480	complement(1674956..1675823)	transposase (partial)	y2690	2972094..2972414	putative transposase	
7	YPO1668	complement(1898862..1900097)	putative membrane protein	y1829	complement(2014178..2015413)	putative resistance protein, transport	
	YPO1669	complement(1900110..1900487)	YjgF-family lipoprotein	y1830	complement(2015426..2015833)	hypothetical protein	
	YPO1670	complement(1900484..1901470)	conserved hypothetical protein	y1831	complement(2015800..2016786)	hypothetical	
	YPO1671	1901648..1902325	putative DNA-binding protein	y1832	2016964..2017641	hypothetical	
	YPO1672	1902800..1906606	conserved hypothetical protein	y1834	2018786..2021923	hypothetical	
8	YPO1817	complement(2064772..2065191)	conserved hypothetical protein	y2490	2751668..2752087	hypothetical	
	YPO1818	complement(2065236..2065604)	conserved hypothetical protein	y2489	2751255..2751623	hypothetical	
9	YPO2006	2279186..2279722	hypothetical protein	y2302	complement(2537129..2537785)	hypothetical	
	YPO2007	2279719..2280549	putative esterase	y2301	complement(2536302..2537132)	hypothetical	
	YPO2008	2280557..2280955	putative rhodanese-like protein	y2300	complement(2535896..2536348)	hypothetical	
	YPO2009	complement(2281003..2281512)	putative integrase core domain protein	y2299	2535339..2535848	putative recombinase	
	YPO2009a	complement(2281533..2281844)	conserved hypothetical protein	NC			
10	YPO2084	join(complement(2363019..2364233),	putative phage integrase (pseudogene)	y2227	2455229..2456467	putative transposase	phage
	YPO2087	complement(2366293..2366553)	putative phage excisionase	NC			
	YPO2087a	complement(2366634..2366834)	putative phage protein	NC			
	YPO2088	complement(2366853..2367497)	putative methyltransferase	y2224	2451986..2452630	putative DNA methyltransferase	
	YPO2089	2367596..2368024	putative phage protein	y2223	complement(2451459..2451905)	putative phage nin-region protein	
	YPO2090	2368098..2368703	putative phage protein	y2222	complement(2450780..2451385)	putative phage ninG-like protein	
	YPO2091	2368704..2369093	putative phage antitermination protein	y2220	complement(2450390..2450779)	hypothetical protein	
	YPO2092	2369097..2369315	hypothetical phage protein	y2219	complement(2450268..2450393)	hypothetical	
	YPO2093	2369364..2369927	putative phage protein	y2218	complement(2449556..2450119)	putative phage protein	
	YPO2094	complement(2370012..2370194)	hypothetical phage protein	NC			
	YPO2095	2370351..2370560	hypothetical phage protein	y2217	complement(2448923..2449132)	hypothetical	
	YPO2096	complement(2370557..2370832)	hypothetical phage protein	NC			
	YPO2097	2370955..2371275	putative phage protein	NC			
	YPO2098	2371306..2371818	putative phage lysozyme	y2216	complement(2447665..2447820)	hypothetical	
	YPO2099	2371803..2372261	putative prophage endopeptidase	y2215	complement(2447222..2447680)	putative phage endopeptidase Rz	
	YPO2100	2372723..2373520	phage regulatory protein	y2214	complement(2445963..2446760)	putative phage antirepressor	
	YPO2101	2373977..2374612	hypothetical phage protein	y2213	complement(2444871..2445506)	hypothetical	
	YPO2102	2374643..2375092	hypothetical phage protein	y2212	complement(2444391..2444840)	hypothetical	
	YPO2103	join(2375102..2375656,2376971..2377915)	putative phage terminase (pseudogene)	y2211			
	YPO2106	join(2377915..2378640,2378644..2379300)	putative phage protein (pseudogene)	y2209	complement(2440840..2441568)	hypothetical	
	YPO2108	2379304..2380416	hypothetical phage protein	y2207	complement(2439067..2440179)	hypothetical	
	YPO2109	2380538..2381311	hypothetical phage protein	y2206	complement(2438172..2438945)	hypothetical	
	YPO2110	2381325..2382530	putative phage protein	y2205	complement(2436953..2438158)	hypothetical	
	YPO2111	2382530..2383060	hypothetical phage protein	y2204	complement(2436423..2436980)	hypothetical	
	YPO2112	2383057..2383311	conserved hypothetical phage protein	y2203	complement(2436172..2436426)	hypothetical	
	YPO2113	2383313..2383663	hypothetical phage protein	y2202	complement(2435820..2436170)	hypothetical	
	YPO2114	2383665..2384249	hypothetical phage protein	y2201	complement(2435234..2435818)	hypothetical	
	YPO2115	2384246..2384653	hypothetical phage protein	y2200	complement(2434830..2435237)	hypothetical	
	YPO2116	2384719..2385639	putative phage protein	y2199	complement(2433844..2434764)	hypothetical	
	YPO2117	2385652..2385963	hypothetical phage protein	y2198	complement(2433520..2433831)	hypothetical	
	YPO2118	2386050..2386271	hypothetical phage protein	NC			
	YPO2119	2386272..2389775	putative phage tail protein	y2197	complement(2429708..2433211)	putative tail length tape measure protein	
	YPO2120	2389778..2390119	putative phage protein	y2196	complement(2429364..2429705)	putative phage tail protein	
	YPO2122	2391003..2391755	putative phage protein	y2195	complement(2428438..2429190)	putative phage tail protein	
	YPO2123	2391758..2392468	putative phage minor tail protein	y2194	complement(2427725..2428435)	putative phage tail protein	
	YPO2125	complement(2393440..2393979)	putative phage regulatory protein	y2192	2426322..2426753	hypothetical	
	YPO2126	2394149..2395243	putative phage protein	y2191	complement(2424950..2426044)	putative phage antirepressor	
	YPO2127	2395342..2395893	putative phage-related membrane protein	y2190	complement(2424300..2424851)	hypothetical	
	YPO2128	2395901..2396251	putative phage-related lipoprotein	NC			
	YPO2129	2396307..2396927	putative phage tail assembly protein	y2189	complement(2423266..2423886)	putative phage tail protein	
	YPO2130	complement(2397102..2397413)	hypothetical phage protein	NC			
	YPO2131	2397499..2400702	putative phage host specificity protein	y2188	complement(2419491..2422694)	putative phage tail protein	
	YPO2132	2400702..2401700	hypothetical phage protein	y2187	complement(2418493..2419506)	hypothetical	
	YPO2133	2401717..2402634	hypothetical phage protein	y2186	complement(2417559..2418476)	putative phage tail protein	
	YPO2134	2402645..2403064	putative phage tail fiber assembly protein	y2185	complement(2417129..2417548)	putative phage tail protein	
	YPO2135	2403061..2403216	hypothetical phage protein	NC			
11	YPO2261	2541334..2541933	hypothetical protein	y2103	2317722..2318321	hypothetical	
12	YPO2466	complement(2764300..2764554)	hypothetical protein	NC			
13	YPO2483	complement(2785296..2785700)	hypothetical protein	y1703	1886401..1886805	hypothetical	
	YPO2484	complement(2785842..2786009)	hypothetical protein	NC			
	YPO2485	complement(2786212..2786409)	conserved hypothetical protein	y1702	1885692..1885889	hypothetical	
	YPO2486	complement(2786419..2788452)	conserved hypothetical protein	y1701	1883637..1885682	putative hemagglutinin-like secreted protein	
14	YPO2503	complement(2813712..2814380)	hypothetical protein	y1685	1857736..1858458	hypothetical	
15	YPO2574	complement(2894374..2894928)	putative membrane protein	y1144	complement(1290571..1291125)	hypothetical	
16	YPO3436	complement(3838865..3839959)	hypothetical protein (pseudogene)	y0751	838498..839619	hypothetical	
	YPO3437	complement(3839952..3841862)	conserved hypothetical protein	y0750	836505..838517	hypothetical	
17	YPO3609	4023133..4024503	conserved hypothetical protein	y0265	complement(278802..280172)	rhsD protein	
	YPO3610	4024505..4024990	conserved hypothetical protein	y0264	complement(278315..278800)	hypothetical	
18	YPO3616	4033596..4034390	conserved hypothetical protein	y0254	complement(268916..269710)	hypothetical	
19	YPO3945	4445294..4445668	hypothetical protein	y3883	complement(4323415..4323789)	hypothetical	
	YPO3946	complement(4445759..4446046)	hypothetical protein	y3882	4323037..4323324	hypothetical	
	YPO3947	complement(4446053..4446427)	hypothetical protein	y3881	4322656..4323030	hypothetical	
	YPO3948	complement(4446668..4446949)	hypothetical protein	NC			
20	YPO4018	4533415..4534359	pyridoxal-phosphate dependent protein	y4039	4480139..4481146	cysteine synthase	
21	YPO4028	4543227..4543589	putative regulator	y4049	4490014..4490376	hypothetical	other
	YPO4029	4543628..4544560	IS903 transposase (pseudogene)	y4050	4490548..4491054	putative transposase	
	YPO4031	4544734..4544979	putative regulatory protein	NC			
	YPO4032	complement(4545247..4546554)	hypothetical protein	y4051	complement(4492034..4493506)	hypothetical	
*only in CO92	YPO2271	complement(2554373..2554636)	phage hypothetical protein	**Absent**			phage
	YPO2272	complement(2554658..2555065)	phage hypothetical protein	**Absent**			
	YPO2273	complement(2555165..2555584)	phage hypothetical protein	**Absent**			
	YPO2274	2555780..2556832	putative phage protein	**Absent**			
	YPO2275	2556844..2557161	putative phage protein	**Absent**			
	YPO2276	2557164..2557415	putative phage-related membrane protein	**Absent**			
	YPO2276a	2557412..2557681	putative phage-related membrane protein	**Absent**			
	YPO2277	2557749..2559461	phage hypothetical protein	**Absent**			
	YPO2278	2559474..2559761	putative phage-related membrane protein	**Absent**			
	YPO2279	2559765..2560790	putative phage-related membrane protein	**Absent**			
	YPO2280	2560790..2562034	putative phage-related secreted protein	**Absent**			
	YPO2281	2562517..2562765	phage hypothetical protein	**Absent**			

NC: present, but not called

Table 6a: Identification of *Y. pestis*-specific genes across a panel of strains of the two species

			Y. pestis strains (O, Orientalis; M, Medievalis; A, Antiqua)													*Y. pseudotuberculosis strains* (Serovars I through V)									Oligonucleotides used for PCR	
	KIM10+	CO92	O	O	O	O	O	M	M	M	M	A	A	A	A	I	I	I	I	II	V	III	III	IV		
Regions from ST3	genes	genes	CO92	297RR	Hambourg10	Exu184	6/69	KIM10+	PKH-4	PKR292	PAR13	Harbin	Japan	Margaret	343	IP32953	IP33134	IP32790	IP32950	IP30215	IP32952	IP32802	IP32889	IP31833	Primer 1	Primer 2
1 - also identified by Hinchliffe et al. (2003)	y3797	YPO0387	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	GCTGCAGGAAGGCAAAGAAGAG	ACCAAAGGTGGTCTGGCTCTTG
	y3796	YPO0388	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CCTCGAAGATGCCGCTGATG	GATGCGAACATCACCGACAGG
	NC	YPO0389	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	TGTTTGTGGGTACAAGTGCG	AGGTAAGGCCTTGCTGTGAC
	y3794	YPO0390	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CACACATCAACACCCAGCTTCC	ATCAGTATTCTCCGCGAGGCTG
	y3792	YPO0391	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	GCTATTCAGGACGAAAGTTCTAAGTGG	GTGGAGAACTTACCAGAAGATCACATTG
	y3791	YPO0392	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CGATTACGGAAACGGTATGTCGG	GCTGCTAATTTCCGCGCAATC
	y3790	YPO0393	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	ACAAGCCTGCATCAATATCCATTG	AACCCACATGTATCCGCACAAAC
	y3789	YPO0394	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	GAAGACACGGCAGAGGCCAAAG	CGTCACTTTCTCCGGCCTTTC
	y3786	YPO0396	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	ATTGCAAGGCCGACTCAATGAG	TAATCCGACCAACGCTTGCAG
	y3785	YPO0397	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	AGCAACTGGTAGCGCAGGTTG	AGCCAACGGAGTGACCAATACC
2	NC	YPO0523	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	+	TTTGAATATTAGCTCGCTGGC	CTGTATTTATACCAAACAAGGGTATTC
3 - assigned as "unique" by Hinchliffe et al. (2003)	NC	YPO1087	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	CTCCGTGATTCGATCGCTAAGG	AAGCAACAACACGACTTCACCG
	y3089	YPO1088	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	AAGCCCGTTTACATGATGACGC	GCGGTAATCGATGCAAGATGTG
	y3088	YPO1089	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	TCAGGATGTCACCGAGCAAACC	CAAATTGAGCGGCGAGAGGTC
	y3087	YPO1090	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	AGGCCGCCAGAGAAATCTCC	CGGGATAACGTTCACGCAGTG
	y3086	YPO1091	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	CGGGCGCTTTGCAGTATAATCC	GAGCATTGCCAATACCAGTGCG
	y3085	YPO1092	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	CATAACGAAATGGCGGCATTG	TGACTGCCCGTTTGACTGCTTC
	y3084	YPO1092a	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	AGCACAAACCTGCCTCAGCC	CGGGCAGCAGTTGAGGTATTG
	y3083	YPO1094	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	AAACCGGTGGTCTACCATGACG	TTAGCACGCGCCAGAGATTGAG
	NC	YPO1095	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	TCGACAGCGAGTTTACATGG	CAGGTTAGAAAGGGATGTAGCC
	y3083	YPO1096	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	TGGGCTAAACCTTGGGATAGTTTC	CTGCGAAGCTTTGCTTTATTTCAC
	y3082	YPO1097	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	CACATCTGACAATCCCTTGTCCG	AAACGTTCTTCCCTCTGATGCG
	y3081	YPO1098	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	GCCAACCCGCTGCTTACTTG	AGCGCAGCCTCAATCACATCAG
4	y2934	YPO1252	+	+	+	+	+	+	+	-	+	-	+	+	-	-	-	-	-	-	-	-	-	-	TATTCGTGGCTGGGATGATGG	ATCATCCGTAACGCCTACTGCC
5	y2929	YPO1257	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	+	+	-	-	AGCCGTGTTTCCGTCTTCACTC	CGGATCTCTGCCAATATAGCGG
	y2928	YPO1258	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	+	+	-	-	TCGTGACCGTGCGGAACTTATC	AATCACACCTTTCGGAACAGGC
6	y2696	YPO1474	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	+	+	-	+	AAGCAACTCGACTGAAGAAGATACG	TTCCAGACTCCAGAGCCTGTAGG
	y2694	YPO1476	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	+	TCAACGAACACGCGCAATAATC	AATTAGTTCGCTGCCTTGCTCC
	y2692	YPO1477	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	+	+	-	+	ACAAAGTATTCCGGCCCATCAATG	TCCACACAAAGCATAGATGTCGTG
	y2690	YPO1480	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	+	+	-	+	CTCTCATACCGCCAGAACGCTC	GATGGTTTCAGCTCGCCGATAG
7 - also identified by Hinchliffe et al. (2003)	y1829	YPO1668	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CGCTTCTGGTGCCCTGTTTG	TGATCCCGTAAGCGCCAATAAG
	y1830	YPO1669	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	-	+	+	-	AACGCACCAGCAGCTATCGG	ACATGAACGGGCAGGGAATG
	y1831	YPO1670	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	AAAGCCGTGCAATTTGCATCTG	GCAGAGGCATAAGGAAAGCAGG
	y1832	YPO1671	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CATCGTTCCGTGATCCACATTG	ACCGGTTCAAGGTGGCTTTGG
8	y2490	YPO1817	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	+	+	-	+	GGCTCGGCAAGCGGTTATTTC	AATTGCTTCGGCTGGCATACTC
9	y2302	YPO2006	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	+	-	-	AGCATGATTGCTCGGTACACGG	TCCCAGCAGGGTTTGTGCAG
	y2301	YPO2007	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	+	-	-	AAATTGCGCAGTGCTGAGCC	TGATGCTAGCGGGTCAAACTGC
	y2300	YPO2008	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	+	-	-	GTTGTGGTCGATGGACGCAG	GAATAACCATCGCGTTTCCACC
	y2299	YPO2009	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	+	-	-	TAGCCGACAGGCGTGGTATCAG	ATGAGGACGCCGGGTGTTGTAG
	NC	YPO2009a	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	+	-	-	GCGCAAGACCGTTACGGTATTC	TGAGTCCTTGCGTTTGTGCTTG
10 - entire region assigned as "unique" by Hinchliffe et al. (2003)	y2227	YPO2084	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	AATCTGTGCAGGGATCATTGGG	TGCCATTTGCTCCTTCAGTGC
	NC	YPO2087	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CCCGCCGAAAGTCATGGATG	GCAGAAATCCTCTCGCCAATCC
	NC	YPO2087a	+	-	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	TTGACTGGATCCTGCTTGG	AATGGCTTAACTTGAGATGCG
	y2224	YPO2088	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CGTTCTGCGCGATATGACACTG	GGGATGAGATAACGCGGGTGTC
	y2223	YPO2089	+	-	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CAAGAGCGAACCCGCACATTAG	TGCTGCCACCTTTCCCTTGTC
	y2222	YPO2090	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	ACGAGGTTCAAGCCAGAGACCG	TGGCGATGGCAATTATCTTCG
	y2220	YPO2091	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	AGCCGAACTCAATGTTGTGACG	AAGCTGCTAACTCCCGCCTGAC
	y2219	YPO2092	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	GCACCCTGGATATCACAACAACG	CATTGGGTGGTGAGACGTACGAG
	y2218	YPO2093	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	TCCTTTCGCAACACTCAATTCG	ATAACGCGCACACGATAACGAG
	NC	YPO2094	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	AACGGATAAGGATGTGGTGC	GAACCTGAATCTTGCCATCC
	y2217	YPO2095	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	AAATTTGCAGAATTACCACATGAAG	TCTTGACCCAGCTTAAGTTTATCG
	NC	YPO2097	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	AACTGACGAGTGGCTGCACTGG	CATTAGCCTTCTCCCGCAACTG
	y2216	YPO2098	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	+	-	-	-	+	TGATCGGCAATGCTGAATCATG	ATCCAGCGCGGGAACTGTTC
	y2215	YPO2099	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	-	-	-	-	CTGTAATTGCCGCCCTGCTTG	CCGCTTATTTGGCTTGTGGC
	y2214	YPO2100	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CCGAGCGGTTAAGGGCTACATG	AAGCCTCACGGAAGCCAAATG
	y2213	YPO2101	+	+	+	+	+	+	+	+	-	+	+	+	+	-	-	-	-	-	-	-	-	-	GTTCTTGGTCAAGCCGAGCAGG	ACCGTATCTGCGAACCACACCC
	y2212	YPO2102	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	ACCTTGTCGATGGTGGCTGAAG	AAGCATCACCAACCTCTTTCGC
	y2211	YPO2103	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	GGCCGAAGCATTCCAACAGTC	CCGCGCCAGTTCTTTACGAG
	y2209	YPO2106	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	GTCGTGCTCCGTGAGTGTGTTG	CGCGCCACTCTTCGGTTAATC
	y2207	YPO2108	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	CTGTTGGACTTCGCAAAGCATG	CGGCAGCCATATGAGTGACAGC
	y2206	YPO2109	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	+	GCTCAAAGAGCAAGGCGAGAAC	CATCACCAAGCACCCGATCAC
	y2205	YPO2110	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	CGGTATTCAGGCGTTGAGTGC	CGTCGATTGGCGCAGTATCAG
	y2204	YPO2111	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	TTTAACGGCGTTTGCTTTAGCG	AATAGCTTCCACGGCCAGCATG
	y2203	YPO2112	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	TTGCCGGATACGACAAAGCTG	CAATCGCGCCAACCATTACG
	y2202	YPO2113	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	ATGCGAGCGACAGCAACACG	CAGTAATGCCGCTGGCTTAACG
	y2201	YPO2114	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	TACGATGTCGCCGGTTGGTAAC	ATATTGCCGTTCGGGTCTTTCC
	y2200	YPO2115	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	AGGTATTCCGTTGGCTGCTGAG	AAATGCCGTTGGTTATGCTTGG
	y2199	YPO2116	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	ATCCCGCAAATCACCGAAGTG	GAATACCACCGCCACTGTTTCG
	y2198	YPO2117	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	-	-	-	-	-	ACATTCACGTTCGACACATGCC	TTCGGGTAGTTATTCGCAAGCAC
	NC	YPO2118	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	CGGATGTCTGGCCTGCTTTC	ATAACGGACAGCGCAACGCTC
	y2197	YPO2119	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	CACTTGGCGCATCTTCAATTGC	CTTGAGCATCAATCGCATCCTG
	y2196	YPO2120	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	+	-	CATTTCTTTGGCGAACTCAGGG	CAACGGGTACGGCCTTGATG
	y2195	YPO2122	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	+	-	AGCCGCCAATGGTGATGAATC	AATCGCAGCCATTACCTGAACG
	y2194	YPO2123	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	+	-	GCGAGAACCTGGCGACAAAC	CAGCATGCCATCTTCCAGCAG
	y2192	YPO2125	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	TTGTCACACCGTGCGTCATTG	AATTTCGGCAAACGTTCTCGTG
	y2191	YPO2126	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	AATGAGCCATGGTTCGTTGCTG	TCCGTGAACTGGTCTTCGCTG
	y2190	YPO2127	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	GCGATCATTTGGCCTATGGTG	GGGTGCATACCGGGATTTATTC
	NC	YPO2128	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	TTTGGTGCTGGCTGGATGC	TGCGACTTCTGTTTGATTTGGG
	y2189	YPO2129	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	TCGGCAAGGATGAGTTGGAGC	ATACCCGCCGAGATAACCGC
	NC	YPO2130	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	GTGTAAGGTCAGCCCACCCAAG	CTGTTTGACCAGCCGCAGTTC
	y2188	YPO2131	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	TCGTCGCTACGGTGTCAATCAG	CAACTGGCATTGGAGGATTTCC
	y2187	YPO2132	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	TACCACAATCACCCGCAGTTGG	CGTTTGGCTGGTTGCTTAGCG
	y2186	YPO2133	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	-	-	CAGGCACAGTCACGTCAGAAGC	AATATTCACATCGCCGGTACCC
	y2185	YPO2134	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	+	-	AATGGACTGGTGGCGTGTGG	ATATCCGGTGCCGCACTTACATC
	NC	YPO2135	+	+	+	+	+	+	+	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	CCGAGGTGGTACGGATTTAG	CAACCTGAAATCATCCAGGC
11 - also identified by Hinchliffe et al. (2003)	y2103	YPO2261	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	CATTTGATTGCTGACCCAACTGG	TCCACACCGCAATCATATATCCC

12	NC	YPO2466	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	+	+	-	-	+	TTGCATTGATTAGGGTGATTGTTTG	CGTGGTTACCATCGGCTTTATACAG
13	y1702	YPO2485	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	TGGTGAATATGTTGGATGGG	TCAGCTAATTGATTATCTAGGAATTG
	y1701	YPO2486	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	TGCGGGTGGGAATGTGTTG	CATCCGCGCTATGGTAAACACC
14 - assigned as "unique" by Hinchliffe et al. (2003)	y1685	YPO2503	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	+	+	-	-	-	TGTCAGGTATTCAATCGGCATCAC	TGCCTTTATGTCGCCATACACAG
15	y1144	YPO2574	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	TGCATGGGCCACGATCATC	TTTCACCCGCCAAGTAAGCTG
16	y0751	YPO3436	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	-	-	-	-	GGATGACTTTGCGCTATACCCG	ATGACGTCCTGCACTGAGGCAC
	y0750	YPO3437	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	-	-	-	-	TCGAGATTGCCACGGATAACG	GTAGCTGTGACCGCCAGTTTGC
17	y0265	YPO3609	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	AATGCAAGGTGCCCGAGGAC	GGAATTCGGCCCTTTAACCAAC
	y0264	YPO3610	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	GTGGTGGGCTTGTGATGATGG	CGTCTTCTTCGACCAATCCATTG
18	y0254	YPO3616	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	GCCTCATTTGGGTTCTCAACAAG	TCACATTAACCGGGTACACCATCC
19	y3883	YPO3945	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	+	-	+	-	AACGGAGTGCTCGCCTGATTG	ACCACCTGCCTTTCCTGTTTCC
	y3882	YPO3946	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-	+	+	+	+	-	GCCATCCATCTTGGAACTCTGG	CCTCTGCGAGGTCGGTAAATTC
	y3881	YPO3947	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-	+	+	+	+	-	TGAAGCATTCCACGGTAAACCATAC	TTGATGCTGTAAGGCAAGATTCTGC
	NC	YPO3948	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	TTTCTAGGCCAACTCATCGGTTG	AAATCATATGCGGCTGTTTCCC
20	y4039	YPO4018	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	-	-	-	-	+	-	-	AGCCTACCCACGAATCCATACG	CTCCACATGTTGGCCCAATTC
21 - assigned as "unique" by Hinchliffe et al. (2003)	y4049	YPO4028	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+	GCAGGAAAGCCTCAGTTTGTGG	TTTCTTCTGCGGGCGAGAGTC
	y4050	YPO4029	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	TCCGTTGCCTGGATTATTCCTG	CATTGGCTGAAGCGACCCTG
	NC	YPO4031	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	CGCCAGTCGCTATTTATCAGAGC	AATGAATTCGTGCGTTAGAACCG
	y4051	YPO4032	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	TTCCCGCTATTAGCTCATGGATTG	AGGCTAACCAATCTTTCAAGTGCAG
*only in CO92	absent	YPO2271	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	AACTTAGGTTATCACATTCGTTCTC	GCATATGCTAGGGCTTCTCTAAG
	absent	YPO2272	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	CCTGCTTTCCGCAAAGAATCTC	TGCTTCCTTCCGGAGGGTTG
	absent	YPO2273	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	TGGCGGACAAACAGAGAGCAG	CAGAAACCTGCCTTGCGAGTG
	absent	YPO2274	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	CCATCCTATCGGCAGGTCATG	TTGGGCGTCTCAATATTCCAGC
	absent	YPO2275	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	TGGTGCTCGCCAGCGTAATAAG	TCAGTAATGCCGCCATCTGATG
	absent	YPO2276	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	TGGATACATCAGGTTTCGATATTC	CCAGAGCAGCAGGACAGAG
	absent	YPO2276a	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	GCAGTTTCTGGCGCGACTTTC	CCAGCGACACCGGCATAAAG
	absent	YPO2277	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	-	GGCCGCATCTTCATTTCTCG	AGCCTAAGTTCCGACATCGCAG
	absent	YPO2278	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	GTAGTCCAGGGCTTCGTTGAAATC	CGAGTAGCCATTGCGGAGATAATC
	absent	YPO2279	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	AAGCATGGCGAATCTTCGGTAG	AAACATGGCCGCATCACGTAC
	absent	YPO2280	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	AGCCGTGATTTCTGACGATTCG	ACCGCTGGAACGATATTCAACG
	absent	YPO2281	+	+	+	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	TTCATGACACATCAGGAGCTGAAAG	AGGCCTGTTCATCCAGTTAGTTGG

NC : present, but not called

Identified in this study as *Y. pestis* -specific genes (or gene segments)

+ : amplification of expected size

- : no amplification

Table 6b: *Y. pestis* -specific gene clusters identified by genome comparison and strain					
***Y. pestis* -specific regions**	**Locations in *Y. pestis* -CO92**		**Locations in *Y. pestis* -KIM10+**		**Product**
		YPO0387		y3797	hyp. protein
		YPO0388		y3796	cons. hyp. protein
		YPO0389		NC	hyp. protein
		YPO0390		y3794	hyp. protein
		YPO0391		y3792	modification methylase
		YPO0392		y3791	hyp. protein
		YPO0393		y3790	hyp. protein
		YPO0394		y3789	hyp. protein
		YPO0396		y3786	hyp. protein
Cluster 1 (11846bp)*	403002-414847	YPO0397	4205743-4217588	y3785	hyp. protein
		YPO1668		y1829	Putative membrane protein
		YPO1670		y1831	cons. hyp. protein
Cluster 2 (8174bp)*	1898559-1906732	YPO1671	2013875-2022049	y1832	Putative DNA-binding protein
		YPO2084		y2227	Putative phage integrase (pseudogene)
		YPO2087		NC	Putative phage excisionase
		YPO2088		y2224	Putative methyltransferase
		YPO2090		y2222	Putative phage protein
		YPO2091		y2220	Putative phage antitermination protein
		YPO2092		y2219	hyp. phage protein
		YPO2093		y2218	Putative phage protein
		YPO2094		NC	hyp. phage protein
		YPO2095		y2217	hyp. phage protein
		YPO2097		NC	Putative phage protein
		YPO2100		y2214	phage regulatory protein
		YPO2102		y2212	hyp. phage protein
		YPO2103		y2211	Putative phage terminase (pseudogene)
Cluster 3 (12666bp - putative lambdoid bacteriophage)* (585bp)	2362991-2375656	YPO2114	2443827-2456492	y2201	hyp. phage protein
Cluster 4 (775bp)*	2541157-2541931	YPO2261	2317545-2318319	y2103	hyp. protein
		YPO2485		y1702	cons. hyp. protein
Cluster 5 (1868bp)	2785156-2787023	YPO2486	1885078-1886945	y1701	cons. hyp. protein
		YPO3609		y0265	cons. hyp. protein
Cluster 6 (1086bp)	4024090-4025175	YPO3610	278130-279215	y0264	cons. hyp. protein

NC = Genes present but Not Called in that genome

* Regions that correspond to those found by microarray analysis (Hinchliffe et al. 2003)

Table 7: Pseudogenes in *Y. pseudotuberculosis* IP32953

Y. pseudotuberculosis IP32953 gene	*Y. pestis* KIM10+ gene	*Y. pestis* CO92 gene	Type of mutation	Product
YPTB0014	y3815	YPO0013	F/S	molybdopterin-guanine dinucleotide biosynthesis protein A
YPTB0146	y0347	YPO3886	partial	colicin
YPTB0148	NC	YPO3884	partial	colicin
YPTB0187	absent	absent	IS285	hypothetical
YPTB0361	y0564	YPO0304	F/S	putative resistance protein (transport)
YPTB0390	y0595	YPO0336	partial	insertion element protein
YPTB0394	NC	YPO0340	partial	transposase
YPTB0534	y0749	YPO3438	stop	Integrase (intB)
YPTB0665	absent	absent	partial	putative cytoplasmic protein
YPTB1018	y1088	YPO3090	F/S	5'-nucleotidase/UDP-sugar diphosphatase
YPTB1042	NC	YPO2819	partial	phage integrase
YPTB1060	NC	YPO2797	multiple	hypothetical
YPTB1129	y1213	YPO2639	partial	integrase
YPTB1131	y1215	YPO2641	partial	phage family integrase
YPTB1134	y3081	absent	partial	putative prophage integrase
YPTB1271	y2951	YPO1236	stop	putative class II aldolase-family protein
YPTB1295	y2926	YPO1259	partial	putative bacteriophage protein
YPTB1364	y2847	YPO1334	inframe deletion	putrescine transport system permease protein
YPTB1466	y2722	YPO1448	partial	transposase
YPTB1468	y2719	YPO1449	multiple	cytotoxic necrotizing factor
YPTB1602	y2393	YPO1917	F/S	integrase
YPTB1707	NC	YPO1833	F/S	conserved hypothetical
YPTB1878	y2439	YPO1869	partial	phage replicative DNA helicase
YPTB1880	y2440	YPO1868	F/S	putative transposase
YPTB1893	y2435	YPO1874	F/S	conserved hypothetical
YPTB1894	y2434	YPO1875	partial	putative major tail sheath protein
YPTB1903	y2426	YPO1883	IS285	hypothetical protein
YPTB1904	y2425	YPO1884	F/S	transposase
YPTB1908	NC	YPO1888	partial	putative transposase
YPTB1909	NC	YPO1889	partial	putative transposase
YPTB2201	absent	absent	partial	na+/H+ antiporter
YPTB2202	y2113	YPO2283	insertion	putative LacI-family transcriptional regulator
YPTB2384	y1859	YPO1697	F/S	pili assembly chaperone protein
YPTB2408	y1822	YPO1663	F/S	flagellin biosynthesis txn reg.
YPTB2495	absent	absent	deletions	glucans biosynthesis protein
YPTB2524	y1701	YPO2486	partial	hemalglutinin-like secreted protein
YPTB2730	y1471	YPO3010	stop	hypothetical
YPTB2748	y1452	YPO3031	partial	putative acetyltransferase
YPTB2752	absent	absent	multiple	integrase
YPTB2930	y1252	YPO2679	F/S	PTS system, cellobiose-specific IIC component
YPTB3040	y3179	NC	partial	Conserved hypothetical protein
YPTB3096	y3236	YPO0851	partial	PTS transport protein
YPTB3127	y3268	YPO0886	partial	insertion sequence protein
YPTB3128	y3270	NC	partial	putative prophage protein
YPTB3178	y3288	YPO0902	F/S	Putative surface protein
YPTB3233	NC	YPO0961	partial	integrase
YPTB3234	y3349	YPO0962	partial	conserved hypothetical
YPTB3240	y3356	YPO0967A	partial	insertion sequence protein
YPTB3253	NC	YPO0979	partial	insertion element protein
YPTB3254	NC	YPO0980A	partial	insertion element protein
YPTB3359	y3476	YPO0702	stop	putative exported protein
YPTB3367	absent	absent	partial	transposase
YPTB3423	y3545	YPO0632	F/S	hypothetical protein
YPTB3450	y3575	YPO0603	partial	putative hemagglutinin/hemolysin-related protein
YPTB3540	y0170	YPO3693	promoter deletion	Conserved hypothetical protein
YPTB3651	y0022	YPO3720	F/S	hemolysin activator protein
YPTB3853	y4039	YPO4018	partial	pyridoxal-phosphate dependent protein
YPTB3854	y4040	YPO4019	start codon mutation	putative phosphoribosyl transferase protein
YPTB3861	y4047	YPO4026	partial	phage hypothetical protein
YPTB3874 and YPTB3877	absent	absent	IS100	mrr restriction system protein, pseudogene
YPTB3927	y4078	YPO4058	stop	formate dehydrogenase-O, major subunit
YPTB3933	y4071	YPO4051	partial	transposase

NC = present but not called
F/S = frameshift
not a pseudogene
a pseudogene or absent from the genome
a pseudogene but previously called as a functional; or missed in the annotation

Table 8a: List of new putative *Y. pestis* CO92 pseudogenes

Y. pseudotuberculosis IP32953 gene	*Y. pestis* KIM10+ gene	*Y. pestis* CO92 gene	Type of mutation	Product
YPTB0039	y0099	YPO0042	F/S	putative membrane protein
YPTB0142	y0343	YPO3891 and YPO3892	F/S	conserved hypothetical protein
YPTB0188	y0383	YPO3847	F/S	frataxin-like protein
YPTB0195	y0390	YPO3840	insertion	putative TetR-family regulatory protein
YPTB0196	y0391	YPO3839	inframe insertion	conserved hypothetical protein
YPTB0207	y0402	YPO3828	F/S	hypothetical
YPTB0221	y0416	YPO3814	inframe deletion	cell division protein
YPTB0243	y0439	YPO3791	F/S	putative membrane protein
YPTB0275	y0474	YPO3757	IS100 in promoter*	putative acetyltransferase
YPTB0315	y0517	YPO0260	F/S	putative AraC-family regulatory protein
YPTB0341	y0544	NC	F/S	HEMIN UPTAKE PROTEIN HEMP
YPTB0362	y0565	YPO0306	F/S	conserved hypothetical protein
YPTB0365	y0567	YPO0309	inframe insertion	ATP-binding component of transferase system
YPTB0368	y0570	YPO0312	inframe deletion	glycerol-3-phosphate acyltransferase
YPTB0389	y0594	YPO0335	inframe deletion	hypothetical protein
YPTB0427	y0633	YPO0375	inframe deletion	putative membrane protein
YPTB0453	y0661	YPO3523	inframe deletion	putative exported protein *YtfN*
YPTB0480	y0688	YPO3496	inframe deletion	translation initiation factor IF2-2
YPTB0545	y3776	YPO0405	inframe insertion	General PTS family, enzyme I, phosphohistidine domain
YPTB0630	NC	YPO0488	F/S	hypothetical protein
YPTB0649	y3665	YPO0508 and YPO0509	stop	Conserved hypothetical protein
YPTB0656	y3658	YPO0515	F/S	putative membrane protein
YPTB0672	y3645	YPO0533	IS1541 in promoter*	2-isopropylmalate synthase
YPTB0714	y0768	YPO3418	inframe deletion	pyruvate dehydrogenase, dihydrolipoyltransacetylase component
YPTB0733	y0789	YPO3398	inframe deletion	putative glutamyl t-RNA synthetase
YPTB0735	y0792	YPO3396	deletion	sugar fermentation stimulation protein; transcriptional regulator for maltose metabolism
YPTB0750	y0808	YPO3381	inframe deletion	bifunctional multimodular BarA: sensory histidine kinase, activates OmpR
YPTB0756	y0815	YPO3375	insertion	superoxide dismutase precursor (Cu-Zn)
YPTB0759	y0818	YPO3372	deletion	sulfite reductase, beta (flavoprotein) subunit
YPTB0801	y0859	YPO3330	IS1541 in promoter*	putative sugar ABC transporter
YPTB0815	y0874	YPO3315	inframe insertion	putative sugar ABC transporter, permease protein
YPTB0824	y0882	YPO3306	Promoter deletion*	putative regulatory protein
YPTB0871	y0939	YPO3249	inframe insertion	putative N-carbamyl-L-amino acid amidohydrolase
YPTB0880	y0942 and y0943	YPO3247	F/S and deletion	putative adhesin; involved in the adherence of host cells
YPTB0932	y0997 and y0998	YPO3186 and YPO3185	F/S	Conserved hypothetical protein ; putative membrane protein
YPTB0977	y1045	YPO3139	inframe deletion	putative 6-O-methylguanine DNA methyltransferase family protein
YPTB1051	y1123	YPO2807	IS100 in promoter*	LysR-family transcriptional regulator
YPTB1064	y1136	YPO2793	Mutation of STOP codon	putative membrane protein
YPTB1066	y1139	YPO2788	F/S	hypothetical
YPTB1087	y1165	YPO2594	F/S	hypothetical
YPTB1121	y1205	YPO2630	IS1541 in promoter*	glutaminyl-tRNA synthetase
YPTB1181	y3032	YPO1150	inframe deletion	adenosylmethionine-8-amino-7-oxononanoate aminotransferase
YPTB1198	absent	YPO1168	inframe deletion	high-affinity choline transport protein
YPTB1256	y2972	YPO1216	inframe deletion	DNA gyrase subunit A
YPTB1271	y2951	YPO1236	partial deletion	putative class II aldolase-family protein
YPTB1286	y2935 and y2936	YPO1250a and YPO1251	F/S	putative bacteriophage protein
YPTB1308	y2909	YPO1274	F/S	putative membrane protein with Diguanylate cyclase/ phosphodiesterase domain 2
YPTB1346	y2870	YPO1315	stop	putative hydrolase
YPTB1370	y2839 and y2840	YPO1344	F/S	FecCD transport family protein
YPTB1401	y2800	YPO1376	inframe deletion	putative cell division protein
YPTB1422	y2775	YPO1397	inframe deletion	conserved hypothetical protein
YPTB1446	y2746	YPO1422	inframe deletion	putative lipoprotein protein
YPTB1458	y2730	YPO1440	inframe deletion	putative helicase IV
YPTB1533	y2650	YPO1518	F/S	putative membrane protein
YPTB1565	y2616	YPO1552	F/S	putative exported protein
YPTB1590	y2405 and y2406	YPO1905a and YPO1905	F/S	hypothetical protein
YPTB1635	y2549	YPO1759	F/S	putative membrane protein
YPTB1639	y2546	YPO1763	inframe deletion	5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase
YPTB1676	y2506	YPO1803	partial deletion	flagellar basal-body rod protein FlgF
YPTB1677	y2505	YPO1804	partial deletion	flagellar basal-body rod protein FlgG
YPTB1700	y2479	YPO1827	F/S	flagellum-specific ATP synthase
YPTB1721	y2457	YPO1849	F/S	conserved hypothetical
YPTB1880	y2440	YPO1868	F/S	putative transposase
YPTB1916	y2413 and y2414	YPO1896	F/S	putative binding-protein
YPTB1920	y2389	YPO1921	inframe deletion	probable pili assembly chaperone
YPTB1980	y2324	YPO1987	IS285	hypothetical protein
YPTB2024 and YPTB2025	y2270	YPO2042	fusion	conserved hypothetical AND virulence factor
YPTB2028	y2267	YPO2045	F/S	putative hemolysin
YPTB2075	y2173	YPO2148	inframe insertion	multidrug resistance protein
YPTB2101	y2146	YPO2175	inframe insertion	Hns DNA binding protien
YPTB2115	NC	NC	F/S	conserved hypothetical
YPTB2117	y2037	YPO2193	inframe insertion	TonB
YPTB2165	y2085	YPO2244	F/S	Fe-S binding NADH dehydrog.
YPTB2166	y2086	YPO2245	stop	Fe-S ferredoxin protein
YPTB2188	y2108	YPO2266	stop	multidrug R permease
YPTB2210	y2118	NC	partial deletion	putative taurine transport system ATP-binding
YPTB2211	y2119	YPO2288	deletion and F/S	amn = AMP nucleosidase
YPTB2212	y2121	YPO2289	inframe deletion	SrfA putative virulence factor in Salmonella
YPTB2213	y2122	YPO2290	inframe insertion	SrfB putative virulence factor in Salmonella
YPTB2217	y2127	YPO2294	deletion	ilvN = acetolactate synthase
YPTB2219	y2129	YPO2296 and YPO2297	F/S	putative membrane protein
YPTB2231	y2140	YPO2309	F/S	two-component reg. system
YPTB2233	y2020 and y2019	YPO2312 and YPO2313	F/S	insecticidal toxin TccC
YPTB2294	absent	YPO2380	inframe insertion	insecticidal toxin
YPTB2403	y1835	YPO1673	inframe insertion	membrane protein
YPTB2406	y1826	YPO1665	inframe deletion	motB chemotaxis
YPTB2415	y1817	YPO1654	inframe deletion	LacZ
YPTB2439	y1788 and y1789	YPO1629	F/S	NAGC-like txn regulator
YPTB2446	y1779	YPO1619	stop	hypothetical
YPTB2486	y1741	YPO2449	inframe insertion	LuxR-like transcriptional regulator
YPTB2498	y1734	YPO2455	partial deletion	permease
YPTB2508	y1724	YPO2465	partial deletion	conserved hypothetical
YPTB2527	y1697	YPO2490	inframe deletion	putative hemolysin
YPTB2531	y1693	YPO2494	inframe insertion	BCCT-family transporter
YPTB2540	y1684	YPO2504	partial deletion	hypothetical protein
YPTB2551	y1672	YPO2515	IS100 insertion	putative chemotactic transducer
YPTB2591	NC	YPO2560	partial deletion	putative exported protein
YPTB2604	y1613	YPO2780	F/S	conserved hypothetical
YPTB2659	y1555	YPO2725	F/S	conserved hypothetical
YPTB2661	y1552	YPO2723	IS100 insertion	possible OmpA family
YPTB2731	y1470	YPO3011	F/S	cysteine synthase B
YPTB2778	y1424	YPO3056	F/S	putative acetyltransferase
YPTB2781	y1421	YPO3059	IS1541 in promoter*	phosphoribosylaminoimidazole-succinocarboxamide synthase
YPTB2823	y1375 and y1374	YPO2858	F/S	hypothetical protein
YPTB2826	y1370	YPO2863	IS285	putative membrane protein
YPTB2842	y1352	YPO2880	inframe deletion	putative DNA-binding protein

YPTB2893	y1294	YPO2715	F/S	putative membrane protein
YPTB3032	NC	YPO0782	F/S	Putative membrane protein
YPTB3052	y3191	YPO0803	F/S	hypothetical protein
YPTB3053	y3192	YPO0804	F/S	Putative regulatory membrane protein
YPTB3054	y3194	YPO0805	F/S	putative membrane protein
YPTB3069	y3208	YPO0820	F/S	hypothetical protein
YPTB3104	y3244	YPO0859	F/S	Sugar ABC transporter system, permease.
YPTB3119	y3258 and y3259	YPO0876 and YPO0877	F/S	No significant database hits.
YPTB3219	y3333	YPO0947	inframe insertion	hemolysin-like Ca binding domains, hits to RTX toxins and related Ca2+-binding proteins
YPTB3234	y3349	YPO0962	partial	conserved hypothetical
YPTB3243	y3359	YPO0968a	F/S	hypothetical protein
YPTB3273	y3433	YPO0761	Mutation of START codon	hypothetical protein
YPTB3282	y3432	YPO0762	F/S	Conserved hypothetical protein.
YPTB3296	y3416	YPO0776	insertion	putative siderophore biosynthesis protein
YPTB3297	y3410 and y3412	YPO0777and YPO0778 and YPO1012	IS insertion	Putative siderophore biosysnthesis protein.
YPTB3308	y3389	YPO0999 and YPO0998	stop	Conserved putative membrane protein
YPTB3322	absent	YPO0743	inframe insertion	flagellar hook-length control protein
YPTB3344	y3460	YPO0718	inframe deletion	Conserved hypothetical protein
YPTB3371	y3483	YPO0694	inframe deletion	Putative membrane protein
YPTB3397	y3513	YPO0666	F/S	Icc protein homologue
YPTB3419	y3539 and y3540	YPO0641a and YPO0642	F/S	hypothetical protein weakly homologous to the small subunit of phage terminases
YPTB3422	y3543	YPO0639	IS100 in promoter*	hypothetical protein
YPTB3424	y3547	YPO0631	stop	LysR-family regulatory protein - unknown function.
YPTB3429	NC	YPO0626	F/S	Conserved hypothetical
YPTB3465	y3585	YPO0594	F/S	Conserved hypothetical protein
YPTB3512	y0139	YPO3568	IS1541 in promoter*	Protease
YPTB3534	y0164	YPO3699	IS100 in promoter*	putative exported protein
YPTB3551	y0183	YPO3681	inframe insertion	Insecticidal toxin TcaA
YPTB3553	y0185	YPO3678	inframe insertion	Insecticidal toxin TcaC
YPTB3579	y0216	YPO3651	inframe deletion	Transcriptional regulator
YPTB3605	y0245	YPO3624	insertion	aliphatic sulfonates binding protein
YPTB3621	y0265 and y0266	YPO3608 and YPO3609	stop	putative exported protein
YPTB3624	y0269	YPO3604 and YPO3605	F/S	Conserved hypothetical protein
YPTB3655	y0018	YPO3724	Insertion in promoter*	isocitrate dehydrogenase kinase/phosphatase
YPTB3712	y3975	YPO0193	inframe deletion	peptidyl-prolyl cis-trans isomerase
YPTB3737	y3948	YPO0164	inframe insertion	putative membrane receptor protein
YPTB3743	y3941	YPO0158	inframe deletion	siroheme synthase
YPTB3756	NC	NC	partial only	putative (AJ306977) glutenin HMW subunit 1Ay
YPTB3785	y3888	YPO3940	F/S	glucose-1-phosphate adenylyltransferase
YPTB3789	y3884	YPO3944	inframe deletion	possible Bacterial Ig-like domain (group 1). Putative invasin
YPTB3795	y3877	YPO3952	deletion	conserved hypothetical protein
YPTB3827	y3842	YPO3987	F/S	putative exported protein
YPTB3841	y3826 and y3827	YPO4002	F/S	dipeptide transport system permease protein

* Genes with promoter anomalies were designated pseudogenes if the IS element, insertion or deletion occurred within ~50bp or the ATG start

Table 8b: List of pseudogenes originally called in CO92 (modified from Parkhill et al. 2001)

Y. pseudotuberculosis IP32953 gene	*Y. pestis* KIM10+ gene	*Y. pestis* CO92 gene	Type of mutation	Product
not a gene	NC	YPO2366	IS100	putative membrane protein
absent	y3787	YPO0395	Partial	transposase
absent	y3214	YPO0824A	IS100	transposase
absent	NC	YPO1086a	Partial	phage integrase
absent	y2690	YPO1480	Partial	transposase
absent	y2211	YPO2103	IS285	putative terminase
absent	y2209	YPO2106	F/S	putative phage protein
absent	absent	YPO3125	Fusion	insertion sequence IS100, transposase and ATP-binding fusion protein
absent	y0751	YPO3436	IS100	hypothetical protein
absent	IS1661	YPO3592A	IS100	insertion element IS1661 protein (split from YPO3702 by inversion)
absent	IS1661	YPO3702	IS100	insertion element IS1661 protein (split from YPO3592A by inversion)
absent	y4050	YPO4029	F/S	IS903 transposase
YPTB0058	y0079	YPO0061	F/S	hypothetical protein
YPTB0085	y0048	YPO0089	Partial	GlpX protein
YPTB0086	y0047	YPO0090	Partial	glycerol kinase
YPTB0090	y0286	YPO0098	IS100	ABC-transport protein, ATP-binding component
YPTB0105	y0302	YPO0115	F/S (7G)	cystathionine gamma-synthase
YPTB0146	y0347	YPO3886	Partial	colicin
YPTB0148	NC	YPO3884	Partial	colicin
YPTB0164	y0358	YPO3870	F/S	guanosine-5'-triphosphate,3'-diphosphate pyrophosphatase
YPTB0261	y0455	YPO3775	IS100	deoxyribonuclease TatD
YPTB0269	y0467	YPO3763	IS1541	conserved hypothetical protein
YPTB0305	y0507	YPO0250	IS100	putative hydrolase
YPTB0390	y0595	YPO0336	Partial	insertion element protein
YPTB0394	NC	YPO0340	Partial	transposase
YPTB0436	y0644	YPO3541	IS100	conserved hypothetical protein
YPTB0524	y0738	YPO3448	F/S (7G)	putative extracellular solute-binding protein
YPTB0553	y3765	YPO0414	F/S	Putative transcriptional regulator
YPTB0582	y3742	YPO0437	F/S	thymidine phosphorylase
YPTB0590	y3734	YPO0446	F/S	ABC transporter ATP-binding protein
YPTB0668	y3649	YPO0526	IS100	sugar efflux transporter
YPTB0761	y0820	YPO3370	F/S	phosphoadenosine phosphosulfate reductase
YPTB0792	y0850	YPO3340	Stop codon	putative exogenous ferric siderophore receptor
YPTB0839	y0900	YPO3288	F/S	anaerobic C4-dicarboxylate transporter
YPTB0843	y0907	YPO3281	IS100	P-protein [includes: chorismate mutase and prephenate dehydratase]
YPTB0870	y0938	YPO3250	IS100	putative transaminase
YPTB0915	y0976	YPO3207	IS100	exonuclease SbcC
YPTB0924	y0986	YPO3198	F/S (10G)	Gamma-glutamyltranspeptidase precursor
YPTB0990	y1058	YPO3126	IS100	putative exported protein
YPTB1000	y1069	YPO3114	F/S (11C)	CDP-D-glucose-4,6-dehydratase
YPTB1005	y1075	YPO3108	F/S	putative glycosyltransferase
YPTB1007	y1077	YPO3105	F/S (9T)	putative O-unit polymerase protein
YPTB1009	y1079	YPO3102	F/S	probable GDP-mannose 4,6-dehydratase
YPTB1010	y1080	YPO3100	F/S (6G)	GDP-L-fucose synthetase
YPTB1018	y1088	YPO3090	F/S	UDP-sugar hydrolase
YPTB1042	NC	YPO2819	Partial	phage integrase
YPTB1059	y1133	YPO2798	IS285	conserved hypothetical protein
YPTB1068	y1141	YPO2572	IS285	penicillin-binding protein 1C (split from YPO2784 by inversion)
YPTB1068	y1141	YPO2784	IS285	penicillin-binding protein 1C (split from YPO2572 by inversion)
YPTB1073	y1148	YPO2579	F/S	putative myo-inositol dehydrogenase
YPTB1123	y1207	YPO2632	F/S	N-acetylglucosaminidase
YPTB1129	y1213	YPO2639	Partial	integrase
YPTB1131	y1215	YPO2641	Partial	phage family integrase
YPTB1165	y3049	YPO1130	IS100	phospho-2-dehydro-3-deoxyheptonate aldolase, phe-sensitive
YPTB1168	y3046	YPO1135	F/S (6G)	aldose 1-epimerase
YPTB1202	absent	YPO1172	IS1541	xanthosine permease
YPTB1236	y2994	YPO1195	Stop codon	putative substrate-binding transport protein
YPTB1259	y2968	YPO1219	F/S (7T)	putative two component sensor kinase
YPTB1295	y2926	YPO1259	Partial	putative bacteriophage protein
YPTB1367	y2844	YPO1337	IS100	arginine-binding periplasmic protein 2 precursor
YPTB1449	y2741	YPO1428	IS100	3-oxoacyl-[acyl-carrier-protein] synthase II
YPTB1466	y2722	YPO1448	Partial	transposase

YPTB1468	y2719	YPO1449	Partial	cytotoxic necrotizing factor
YPTB1566	y2614	YPO1554	F/S	sugar transport ATP-binding protein
YPTB1572	y2605	YPO1562	IS285	putative intimin
YPTB1609	y2581	YPO1582	IS285	lactose permease
YPTB1610	y2579	YPO1728	IS100	Putative transcriptional regulator of sugar transport
YPTB1628	y2559	YPO1752	IS100	hypothetical protein
YPTB1637	y2547	YPO1761	F/S	4-hydroxyphenylacetate degradation bifunctional isomerase/decarboxylase
YPTB1668	y2515	YPO1793	IS200-like	invasin
YPTB1709	y2472	YPO1835	F/S	putative chemotaxis protein
YPTB1878	y2439	YPO1869	Partial	phage replicative DNA helicase
YPTB1894	y2434	YPO1875	Partial	putative major tail sheath protein
YPTB1903	y2426	YPO1883	IS285	hypothetical protein
YPTB1904	y2425	YPO1884	F/S	transposase
YPTB1908	NC	YPO1888	Partial	putative transposase
YPTB1909	NC	YPO1889	Partial	putative transposase
YPTB1922	y2387	YPO1923	F/S	Putative sensor protein
YPTB1964	y2344	YPO1967	IS100	porin
YPTB1971	y2335	YPO1977	F/S	3-oxoacyl-[acyl-carrier-protein] synthase III
YPTB2011	y2285	YPO2027	IS100	conserved hypothetical protein
YPTB2016	y2278	YPO2034	IS1541	Putative ABC transporter ATP-binding protein
YPTB2097	y2150	YPO2171	IS285	formyltetrahydrofolate deformylase
YPTB2120	y2041	YPO2197	F/S	putative integral membrane protein
YPTB2169	y2090	YPO2249	F/S	putative toxin transport protein
YPTB2170	y2092	YPO2250	F/S	putative toxin secretion protein
YPTB2204	y2115	YPO2285	Partial	putative ribose-binding periplasmic protein
YPTB2308	y1940	YPO2398	IS100	Putative exported protein
YPTB2321	y1927	YPO2412	Stop codon	Conserved hypothetical protein
YPTB2355	y1888	YPO1726	IS285	pectin degradation protein
YPTB2394	y1846	YPO1684	F/S (9G)	putative membrane protein
YPTB2400	y1838	YPO1676	Stop codon	methyl-accepting chemotaxis protein
YPTB2409	NC	YPO1662	F/S	Flagellar biosynthesis regulatory protein
YPTB2412	y1819	YPO1657	F/S	methyl-accepting chemotaxis protein
YPTB2484	NC	YPO1584	IS285	conserved hypothetical protein
YPTB2601	y1617	YPO2570	IS285	phosphotransferase enzyme II, B component
YPTB2661	y1552	YPO2933	IS100	hypothetical protein
YPTB2668	y1543	YPO2943	IS285	outer membrane usher protein
YPTB2725	y1477	YPO3004	F/S	putative aminopeptidase
YPTB2740	y1460	YPO3021	F/S (8C)	putative outer membrane protein
YPTB2798	y1403	YPO2831	F/S (6A)	putative flagellar associated lysine-N-methylase
YPTB2817	y1382	YPO2851	F/S (7G)	two-component system, sensor kinase
YPTB2826	y1370	YPO2861	IS285	hypothetical protein
YPTB2849	y1345	YPO2887	F/S	putative autotransporter protein
YPTB2871	y1318	YPO2909	F/S (10A)	conserved hypothetical protein
YPTB2910	y1273	YPO2698	Stop codon	conserved hypothetical protein
YPTB2938	y1243	YPO2671	F/S (7G)	urease accessory protein
YPTB2951	y1229	YPO2657	IS285	putative mobilization protein
YPTB2962	y3092	YPO1084	IS100	conserved hypothetical protein
YPTB2962	y1218	YPO2644	IS100	conserved hypothetical protein
YPTB3031	y3170	YPO0781	IS100	putative surface structure protein
YPTB3063	y3202	YPO0814	F/S	general secretion pathway protein F
YPTB3096	y3236	YPO0851	Partial	PTS transport protein
YPTB3127	y3268	YPO0886	Partial	insertion sequence protein
YPTB3178	y3288	YPO0902	Partial	putative surface protein
YPTB3220	y3335	YPO0948	F/S (5A)	conserved hypothetical protein
YPTB3233	NC	YPO0961	Partial	integrase
YPTB3240	y3356	YPO0967A	Partial	insertion sequence protein
YPTB3253	NC	YPO0979	Partial	insertion element protein
YPTB3254	NC	YPO0980A	Partial	insertion element protein
YPTB3263	y3380	YPO0989	F/S	putaive siderophore biosynthesis protein IucA
YPTB3269	y3438	YPO0757	IS1661	hypothetical protein
YPTB3285	y3429	YPO0765	F/S	putative adhesin
YPTB3297	y3406	YPO1012	IS100	putative peptide/polyketide synthase subunit
YPTB3313	absent	YPO0752	F/S (9G)	putative lipoprotein
YPTB3336	y3451	YPO0727	IS100	putative flagellar basal-body rod protein
YPTB3357	y3473	YPO0704	F/S	flagellar assembly protein
YPTB3372	y3484	YPO0692	F/S	putative membrane protein
YPTB3389	y3503	YPO0674	F/S	conserved hypothetical protein
YPTB3423	y3545	YPO0632	F/S	hypothetical protein
YPTB3450	y3575	YPO0603	Partial	putative hemagglutinin/hemolysin-related protein
YPTB3489	y0112	YPO3544	IS100	conserved hypothetical protein
YPTB3527	y0155	YPO3583	Stop codon	probable sigma (54) modulation protein
YPTB3545	y0176	YPO3687	F/S (6G)	putative aldehyde dehydrogenase
YPTB3552	y0184	YPO3679	F/S (6A)	insecticidal toxin complex protein TcaB
YPTB3635	y0281	YPO3593	Partial	hypothetical protein
YPTB3635	y0040	YPO3704	IS166	conserved hypothetical protein
YPTB3654	y0019	YPO3723	F/S (5G+5C)	acetate operon repressor
YPTB3659	y4029	YPO0247	IS100	putative transferase
YPTB3716	y3970	YPO0188	F/S	binding protein dependent ABC-transport protein
YPTB3731	y3954	YPO0170	IS100	acetylornithine aminotransferase
YPTB3775	y3901	YPO0124	F/S	maltodextrin phosphorylase
YPTB3782	y3891	YPO3937	Partial	Aerobic glycerol-3-phosphate dehydrogenase
YPTB3837	y3831	YPO3998	F/S (7T)	endoglucanase
YPTB3846	y3820	YPO4008	IS100	two-component system, sensor kinase
YPTB3861	y4047	YPO4026	Partial	phage hypothetical protein
YPTB3895	y4062	YPO4042	F/S	fimbrial usher protein
YPTB3902	y4104	YPO4087	F/S	conserved hypothetical protein
YPTB3927	y4078	YPO4058	Opal codon	Formate dehydrogenase-O, major subunit
YPTB3932	y4073	YPO4052	F/S (10G)	selenocysteine-specific elongation factor
YPTB3933	y4071	YPO4051	Partial	transposase

F/S = frameshift
NC = present but not called
not a pseudogene
a pseudogene or absent from the genome
a pseudogene but previously called as a functional; or missed in the annotation

Table 9: Pseudogene distribution across a panel *Yersinia pestis* and *Y. pseudotuberculosis* strains

	Y. pestis strains (O, Orientalis; M, Medievalis; A, Antiqua)													*Y. pseudotuberculosis strains* (Serovars I through V)									
	O	**O**	**O**	**O**	**O**	**M**	**M**	**M**	**M**	**A**	**A**	**A**	**A**	**I**	**I**	**I**	**I**	**II**	**V**	**III**	**III**	**IV**	
IP32953 gene	**CO92**	**297RR**	**Hambourg10**	**Exu184**	**6/69**	**KIM10+**	**PKH-4**	**PKR292**	**PAR13**	**Harbin**	**Japan**	**Margaret**	**343**	**IP32953**	**IP33134**	**IP32790**	**IP32950**	**IP30215**	**IP32952**	**IP32802**	**IP32889**	**IP31833**	**Assigned category**
YPTB0105	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	*Y. pestis* -specific
YPTB0261	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB0269	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB0436	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB0524	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB0761	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB0839	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB1059	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB1073	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB1123	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB1202	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB1236	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB1259	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB1668	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2097	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2120	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2169	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2170	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2661	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2668	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2826	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2962	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3269	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3389	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3489	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3527	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3545	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3654	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3837	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3895	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB0915	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	Orientalis-specific
YPTB0990	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB1367	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2400	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3031	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3336	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3782	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB0668	M	M	M	M	M	wt	wt	wt	wt	wt	wt	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	only in Orientalis and some Antiqua
YPTB1068	M	M	M	M	M	wt	wt	wt	wt	wt	wt	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB1165	M	M	M	M	M	wt	wt	wt	wt	wt	wt	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2011	M	M	M	M	M	wt	wt	wt	wt	wt	wt	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2308	M	M	M	M	M	wt	wt	wt	wt	wt	wt	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2951	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2355	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	M	M	wt	M	wt	in some *Y. pseudotuberculosis*
YPTB3178	M	M	M	M	M	M	M	M	M	M	M	M	M	wt	wt	wt	wt	M	wt	M	wt	wt	
YPTB3731	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	CO92-specific
YPTB0090	M	wt	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	not assigned
YPTB0164	M	M	M	M	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB0870	M	M	M	M	M	M	M	M	M	wt	wt	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB2601	M	wt	M	M	wt	M	M	M	M	wt	wt	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3846	M	wt	M	M	M	M	M	M	M	wt	wt	M	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	
YPTB3932	M	M	M	M	M	wt	M	wt	wt	M	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	

Oligonucleotides used for PCR	
Primer 1	**Primer 2**
TGTAACCGGTGCTGCGTTTG	GCAGACACCACCATGAGATCGC
CTGCAACCTGCTGGACGATATG	GCTTGGCAGCACAGACTTTGG
AGCAAGCTGCGCATAGTCACAC	CGATTCTTTGTTGGCTCCTTGG
CACCCTCAATCAATCCAGCCTG	TTGCCTCTTCGCAGCCATTC
AACACCGCTAATGCCGCTAACC	TTTCTGGATCTGCTTCAGCGAC
AACCACACTCGCGTTTCAGACC	GAGCGTGTCAGTTGGGCATTG
CGTTATTAGCGCCCTTCACGAC	ATAACGGCACCGGCGATCAG
TGTAACGGCACAGAGCATCAGC	ACTCACGAAACGGCGTTCAAC
AAACCGGTCCATCTCGATCTTG	AGCGACAAAGTGTTCCCATTCC
ATGATAACGCTGACCAACACCG	GGCCGTTGGCTGAATCACTG
CAGCCTCTCCGGATTTGAATTG	ATCAAAGGCGCAACCATAGACC
GAGCAACCAAAGTCGTACGTGG	GGGATGTTCGGGTGAGCTTTC
TAGGCGTTCATCGGTGAGAATG	ACCGATTTGCAACAACAGCATG
GAGAACGGTTGGCTGTACGGAC	TGTGCGTTCACCTGATAAACCTG
TCACCCGTGAGCAGCATGATC	ACCACTGGCCTTGCTGCTTATC
ATGATGCTCACCGCTCGCAG	TCCACCAAATTCAGGATACGCC
CCTGCGATTAGCGGCAATATC	GTAATTCGTCATGGCTGGCAAG
TTTACGGGCCAAAGCACACAG	GCTCCCTCATGGCTTCAGATTG
TCGCGACAATGAGGAAACCAC	ATTATTGGAAGCGGCACAGCC
TCCCGCCAATAGGTATGCTGAC	ATCCGGCGATTGTCCAAGG
TCCAGTTGGGAAGGCTACATGG	TGAGACCTTGCTGGCCGATTC
TTGCCGTTTCAACGAGTCAGTG	TGACCATTCTCCACCACCAGC
CAAGCTACAGAGGTGCTGGCTG	AGCGCCCAATGCTGTGTACC
ATATGCTGCGCTGCCAGTCC	GTAACCGCCGTTGACGATCC
TAGCATCGACCGCTTACTGGG	AATCGAGGTGGTAAAGTGCATCG
AACCCTGCCAAGCCATTAAGC	TGCAATGCTGGATCGTTGGTC
CGGTAACGCGCCTTACATCG	GCACCGAACCTTCACGACCTAG
GAAACCCAAGGCAGCCACAC	AACCTTGCTTGCGGATCAGC
TTGGGTTGAGACGCCAGGTATC	CCGATGGCACTTGTGATGTCTG
AAGCGGTTGCCGTAATTGAGC	TTGCCGGATACCGTTTCTCG
CGGATGCCATACGCGATACTC	TCCTTGAGGCTTACCCTGTGGG
TGCGATGGGTCAATAACTGGC	GGTTGGCAGTGGTATTGGGAAC
GATGTTATCCGGGCGCACAG	TTTGCAGCCTCAGCCACCTAC
AGCTAAAGGCGGAGAACTGACG	GCGTGGCCAATGATTGCTG
TGCTCAGCGCTATCAGGCTACC	GCAGAGCTGGCATCATTACCG
ATGACCGCCCAGCGTAACTATC	CGGGATGATCGCTATCGTGC
AAGCGCTGTTTCGAACATTCAGAG	ATGCCGGGCGCAGATGAG
CCTCAATGCTGATGTGGACGTG	ACTCATCCTACCGCCAGCTGAC
GCATTGGTGCCGTTAACTTGC	AGCAACTGACGCTGCCAGAAG
TGATTAAAGCGGGCCTGGAAC	CCCAACTGGCGTGCCATTAG
GCGCAGGAAAGCATTCAAGC	CAGTGAATTGGTCGGTCAAGGG
GGGCTGAACCAGCTTAGGCATC	GAAGTTCATCAGCCGTTGTCCC
ACGTCATTTAAGGTTTCTTTACGCC	TGGATGAGGTTATGCTCAGATTTCG
TGGGAAGCGTTAGGCAATGG	AACGTAAATAGCGGCAAATGCG
GGCCTAATGTTCCCGATACCG	ATGGCGGGTGTAACCAACTAGC
CAGCAGGGTGAAACGTTGTGG	GATATCCGGCGTCACACCATG
CAATTTGCTCTTCGGAAGCCTG	GGCCTGACAGTGAATTTACCCG
TTTATGCTGCCATCGATCTTGG	AACTTGCCAGCCATGTTCACG
TGCGCCCTGAACTGTGGTAAAC	CCACAATTCGCCGAACAGATTC
TTAGCCGTCAGTGCTGAAGTCG	TTTGCCGAAACTCTTCTGTGCC
CACCATCATAAACCCGCAATGC	GGCACGTCAAATCTGGTTACCG
TGACATTGCGCCAGATGCTC	CACTGACGAAGCCGGTCATTG

M : Mutant allele

wt : Wild type allele